NOTE TO USERS

This reproduction is the best copy available.

UMI®

Now faith is the substance of things hoped for, the evidence of things not seen.
Hebrews 11:1 (KJV)

University of Alberta

CLASSIFICATION-BASED GLIOMA DIFFUSION MODELING

by

Marianne Morris ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

Department of Computing Science

Edmonton, Alberta
Fall 2005

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-09242-4

Your file Votre référence
ISBN:
Our file Notre reterence
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

To my parents Emad & Mona, and to my brother Samuel, who supported me in many ways.

Abstract

Gliomas are diffuse, invasive brain tumours that originate from a single glial cell and infiltrate through adjacent healthy tissue as the number of tumour cells exponentially increases. The goal of this thesis is to study glioma diffusion and to propose a classification model that would shed some light on glioma growth patterns.

We introduce a 3D classification-based diffusion model, CDM, that predicts how a brain tumour will grow at a voxel-level on the basis of features specific to the patient and the tumour, and attributes of that voxel and its neighbours. We use Machine Learning algorithms to learn this probabilistic general model, based on the observed growth patterns of gliomas from other patients. We demonstrate that our learned CDM model can, in many cases, predict glioma growth more effectively than two standard models: uniform radial growth across all tissue types and another that assumes faster diffusion in white matter. We study CDM numerically and analytically on clinical data.

Acknowledgements

Many thanks to my supervisors Dr. Russell Greiner and Dr. Joerg Sander for their time, encouragement and support throughout this research study. Special thanks to Dr. Albert Murtha, from Radiation Oncology at the Cross Cancer Institute, Edmonton, Alberta, for his time and contribution to this project. Also, thanks to Mark Schmidt for helping in the preprocessing of the data, thanks to Stephen Walsh for his help with the final experiments, and to Chi Hoon Lee, Alden Flatt, Luiza Antonie and Gabi Moise for working on different parts of the project.

Contents

1 Introduction 1
- 1.1 Problem Definition 1
- 1.2 Medical Background 3
  - 1.2.1 Brain Anatomy Overview 3
  - 1.2.2 Glioma Invasion and Difficulties in Treatment 4
- 1.3 Cancer research 6
  - 1.3.1 Mathematical and Computational Modeling 6
  - 1.3.2 Brief Overview of Proposed Approach 7
- 1.4 MR Brain Imaging 8
- 1.5 Radiotherapy 9
- 1.6 Machine Learning 10
- 1.7 Motivations for glioma growth modeling 11
- 1.8 Scope of Thesis 11
- 1.9 Thesis Outline 12

2 Related Work 13
- 2.1 Glioma Treatment 13
- 2.2 Tumour Modeling 15
  - 2.2.1 Macroscopic and Volumetric Modeling 15
    - Simulating Glioma Growth with 3D Cellular Automaton 16
    - Hyperbolastic Tumour Modeling 17
    - Macroscopic Modeling and Supervised Treatment Planning 17
  - 2.2.2 Glioma Modeling Based on White Matter Invasion 18
    - Differential Motility in Grey and White Matter 19
    - White Matter Tract Invasion 20
  - 2.2.3 Statistical Analyses of Glioma Diffusion 21
    - Estimating the Tumour Growth Curve with Linear Regression 21
    - Tracking Growth Rates with Nearest-neighbour & Surface Modeling 22
    - Modeling Mass Effect based on Biomechanical Simulations 22
- 2.3 Discussion 22
- 2.4 Summary 24

3 Glioma Modeling Framework **26**
3.1 Noise Reduction 28
3.1.1 Inter-slice Intensity Variation Reduction 28
3.1.2 Intensity Inhomogeneity Reduction 29
3.1.3 Slice Alignment 29
3.2 Registration 31
3.2.1 Linear Registration 32
3.2.2 Non-linear Warping 33
3.2.3 Spatial Interpolation 33
3.2.4 Summary of the Registration Task 34
3.3 Intensity Standardization 34
3.4 Tumour Segmentation 35
3.5 Temporal Interpolation 37
3.6 Tissue Segmentation 39
3.7 Feature Extraction 40
3.8 Classification 41
3.9 Diffusion Modeling 41
3.10 Known Problems 42
3.11 Summary of the Diffusion Model Framework 43

4 Predicting Glioma Diffusion **45**
4.1 Introduction to Glioma Modeling 45
4.1.1 Background Discussion 45
4.1.2 Motivating the Supervised Learning Task 46
4.1.3 Proposed Approach 46
4.2 Defining the Prediction Task 47
4.2.1 Task Definition 47
4.2.2 Overview of the Prediction Framework 48
4.3 The Training Data 49
4.3.1 Data Noise 49
4.3.2 Class Imbalance 50
4.3.3 Time-interval and Volume Variations 51
4.4 Learning and Classification 53
4.4.1 Task Description 54
4.4.2 The Classifiers 54
4.4.3 Training and Testing Data Folds 56
4.5 The Feature Space 57
4.6 Diffusion Models 62
4.7 Summary 67

5 Results **68**
5.1 Performance Evaluation of CDM 68
5.2 Feature Selection 70
5.3 Experiments and Results Analysis 74
5.3.1 Experiments and Results Comparison 76

5.3.2 Average and Best Case Results 77
5.3.3 Special Cases of Tumour Growth 78
5.3.4 Fairness of Dataset 81
Model Performance versus Tumour Grade 82
Statistical Evaluation of the Three Models using the t-Test 87
5.3.5 Evaluation of Training Features 89
Comparison of Results using Different Feature Sets 89
Comparison of Classifiers' Performance 90
5.3.6 Fairness of Performance Evaluation 91
5.4 Conclusion 92

6 Conclusion **99**
6.1 Contribution 99
6.2 Future Directions 100
6.2.1 Framework Improvements 100
6.2.2 Data-related Suggestions 101
6.2.3 Suggestions for the Classification Task 102
6.2.4 Diffusion Model Improvements 104
6.3 Summary 104

Glossary of Terms **105**

Bibliography **110**

List of Figures

1.1 Illustration of voxel regions involved in training classifiers 3
1.2 Overview of the human brain 4
1.3 Example of glioma recurrence after treatment 6
1.4 MR image modalities 9

2.1 Zizzari's prediction of PTV [138] 18

3.1 Example of inter-slice intensity variation reduction 29
3.2 Example of slice misalignment along the z axis 30
3.3 Reduction of slice misalignment after image registration with SPM 30
3.4 Example of an image volume before and after registration 32
3.5 Spatial priors used in registration and feature extraction 33
3.6 Example images before and after intensity standardization 35
3.7 Example of segmented tumour and edema volumes 36
3.8 Comparison of the real tumour with temporal interpolation results 39
3.9 Example of segmented grey and white matter and CSF 40
3.10 Overview of the framework 44

4.1 Example voxel regions at the slice level, used in training the classifiers . . . 51
4.2 Support Vector Machine classification of a 2D linearly separable dataset . . 57
4.3 AAL and Brodmann anatomical labels 59
4.4 Plotting of image intensities from T1, T1-contrast and T2 modalities 60
4.5 Generic Diffusion Model 64
4.6 Tumour-induced pressure on the ventricles 66

5.1 Patient #9 image volumes and CDM results 78
5.2 Patient #106 image volumes and CDM results 79
5.3 Patient #160 image volumes and CDM results 79
5.4 Patient #165 image volumes and CDM results 80
5.5 Patient #6 image volumes and CDM results 82
5.6 Patient #25 image volumes and CDM prediction results 83
5.7 Patient #37 image volumes and CDM results 84
5.8 Patient #44 image volumes and CDM results 84
5.9 Patient #83 image volumes and CDM prediction 85
5.10 Patient #93 scans and CDM results 85
5.11 Patient #108 scans and CDM results 86

5.12 Patient #147 scans and CDM results 86
5.13 Patient #169 scans and CDM results 87

List of Tables

4.1 Tumour grade and growth rate information for each patient in the dataset . . 63

5.1 Precision and Recall obtained in initial tests for the three models using the IFS feature set (listed in Table 5.6) with 5-fold 'patient-level' cross-validation 71
5.2 Information Gain (IG) scores ranked in descending order for the FS_0 feature set, a subset of the feature space, described in Section 4.5 94
5.3 Experiment I Results 95
5.4 Experiment II Results 96
5.5 Experiment III Results 97
5.6 The four feature subsets used in the experiments 98

Chapter 1

Introduction

Cancer, a notorious disease, has affected and devastated many lives. This disease has been the centre of attention of thousands of researchers for many decades, around the world. Researchers have combined their knowledge and efforts from many areas ranging from medical to mathematical sciences, to better understand the disease and to find more effective treatments. Yet, there is far more to be done before cancer can become a curable disease.

Life-threatening cancer tumours originate from a single cell that has undergone a chromosomal or genetic mutation that affected its protein balance, increasing its capacity for mitotic division or decreasing the suppression mechanism for cell division. As the malignant cell loses its normal function, it starts to abnormally divide much faster than surrounding normal tissue. A small mass of malignant cells results from that single dividing cell, and starts to invade adjacent normal tissue, and to threaten the affected organ.

In this thesis, we use machine learning to study and model the growth of gliomas, brain tumours that originate from a single glial cell of the nervous system.

We first define the problem of glioma diffusion modeling, followed by a section that presents a general overview of the brain anatomy, and of glioma diffusion and treatment. Section 1.3 introduces cancer research and briefly covers computational modeling contributions to glioma growth prediction, including our proposed approach. In Sections 1.4 and 1.5, we describe Magnetic Resonance Imaging and its role in detecting brain tumours, then introduce radiotherapy. In Section 1.6, we give an overview of machine learning contributions and the role of learning and classification in predicting the growth of brain tumours. Section 1.7 motivates the study of glioma diffusion modeling. We conclude this chapter with a summary of the scope and contribution of this research study, and we outline the remaining chapters in this dissertation.

1.1 Problem Definition

The problem that we address in this thesis is the prediction of future glioma growth in patients diagnosed with primary brain tumours. Since there are several interpretations to the problem of tumour growth prediction and diffusion modeling, we provide a formal definition of the research work involved in our task.

In this thesis, we define 'tumour growth modeling' as the task of predicting the size and shape of the tumour at a later time based on attributes of the visible tumour mass (detected on the MRI brain scan) and of the regions adjacent to the tumour. This glioma growth modeling is based on the assumption that occult malignant cells, which have infiltrated through adjacent normal tissue, are responsible for further tumour growth. By 'occult cells', we refer to cancer cells and tendrils that may have infiltrated through normal tissue adjacent to the original tumour but are initially found in low concentrations, below the detection threshold of MR imaging. As these cells build up into small tumour masses over a period of time, they can then be detected on MRI scans. The work in this thesis takes into account the spatio-temporal aspect of the problem of modeling glioma diffusion.

Modeling tumour growth has an application in treatment planning, and may help improve the effectiveness of current treatment modalities. Proper treatment requires knowing where both the visible tumour and the occult cells are. Conventional treatment methods implicitly assume that occult cells are radially spread in all directions alike, but these methods have not been effective in determining the boundaries of the treatment volume. Therefore, an effective diffusion model should be able to accurately 'track' potential tumour growth in regions adjacent to the *Gross Tumour Volume* (GTV). By *GTV*, we refer to the abnormal tumour region visible on the MRI scan.

We present an alternative approach to modeling tumour growth. We use data from a set of patients to *learn* the parameters of a 'general' diffusion model. The data consists mainly of a series of MRI brain scans for patients diagnosed with gliomas. We apply supervised learning to a set of features obtained from the patient data. In particular, given properties of the patient, the tumour and each voxel at one time, our classification-based diffusion system, CDM, predicts the tumour region at a later time.

Our performance system is initialized with a GTV, of a given patient at one time point, and estimates what the tumour will look like once it is some k voxels larger. Our system uses a diffusion model to predict this further tumour growth (*i.e.*, where these k voxels will be). In order to predict these k voxels, our system uses a probabilistic classifier learned by a supervised learning system. The input to our supervised learning system is a time series based on a set of pairs of labeled brain volumes, from various patients, where each pair is from the same patient, but at different times. We include a number of properties about each voxel in each brain volume, as well as a label: *tumour* or *non-tumour*. For each patient (represented by a pair of tumour volumes v_1 and v_2 at different times t_1 and t_2 respectively), we use as training data the voxels in the set difference between this pair of volumes (*i.e.*, $v_2 - v_1$) and that a perfect diffusion algorithm would consider, in growing from the patient's initial tumour v_1 to the final tumour v_2, as well as the border immediately around the tumour volume v_2. Figure 1.1 shows the voxel regions that are used in the training, *i.e.*, voxels that are transitively adjacent to the tumour v_1 up to a 2-voxel border around the tumour v_2. The output of the learner is a probabilistic classifier that, given (a description of) a voxel adjacent to the current tumour volume, returns an estimate of the probability that the voxel will become tumourous.

Model results are evaluated against the *Gross Tumour Volumes* (GTV) manually segmented by radiologists. Manual delineations of *GTV* are based on image abnormalities detected as enhancing intensity regions and abnormal textures, and are limited by the image acquisition protocols and radiologists' subjective interpretations. Therefore, the aim of this work is to predict tumour growth in terms of what human experts define as tumour.

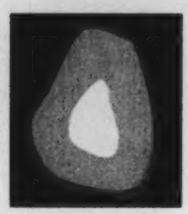

Figure 1.1: Illustration of voxel regions involved in training classifiers
The white region in the centre of the image represents the tumour volume at an initial time point. The light grey region represents additional tumour growth between this initial time point and a later time. The dark grey region consists of normal tissue immediately adjacent to the enlarged tumour. Voxel regions used in the training data are the light and dark grey regions excluding the white area.

The proposed model can help specify the treatment boundaries more precisely. We anticipate that our model's prediction will correspond to regions where radiologically occult cancer cells concentrate but do not enhance on the MRI scan. Therefore, the model can help define the appropriate radiation doses to deliver to the visible tumour and to the adjacent regions, increasing the likelihood of patient survival.

1.2 Medical Background

In this section, we briefly overview some of the brain biology that we need to know for the purpose of this thesis. We then present some information about glioma growth and the challenges in treating malignant brain tumours.

1.2.1 Brain Anatomy Overview

The human brain consists mainly of two tissue types: grey matter and white matter. Grey matter is made of neuronal and glial cells that control brain activity while the cortex is a coat of grey matter that covers the brain. White matter fibre tracts are myelinated axon bundles located throughout the inner regions within the brain, that form pathways between grey matter regions. The corpus callosum is a thick band of white matter fibres connecting the left and right hemispheres of the brain [15, 85]. See Figure 1.2 for a description of each of each of these anatomical regions.

The brain also contains cerebrospinal fluid (CSF) that consists of glucose, salts, enzymes, and white blood cells. The CSF is secreted into the ventricles (Figure 1.2), occupies the space between the skull and the cerebral cortex, and circulates through the ventricular system of the brain and of the spinal cord. Its main function is to cushion the brain against trauma and to help protect the spinal cord from mechanical shocks [134].

Glial cells are the major non-neuronal cell type in the nervous system. Glial cells function as supporting cells that help maintain the signalling ability of the neurons but do not participate directly in electrical signalling [15]. There are three types of glial cells: astrocytes, oligodendrocytes and microglia. The first provides support for the neurons, and help

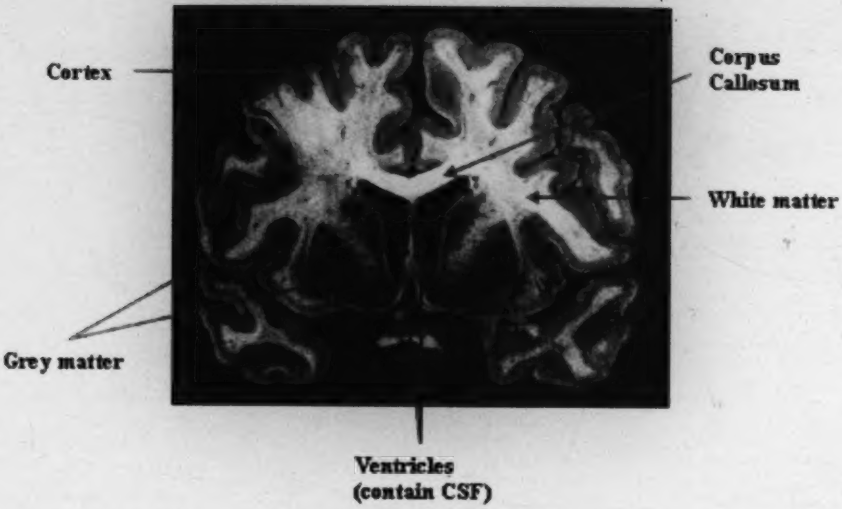

Figure 1.2: Overview of the human brain
Coronal section of the human brain. The cortex is made of grey matter and is connected, by the white matter fibres, to the rest of grey matter in the brain. The corpus callosum is a white matter tract that connects the left and right hemispheres. The cerebrospinal fluid (CSF) is secreted into the ventricles and circulates to the outer surface of the brain and to the spinal cord.
(This image was reproduced from Figure 5-30 acquired from http://www.vh.org/adult/provider/anatomy/BrainAnatomy/BrainAnatomy.html)

regulate the chemical environment by contributing to the blood-brain barrier [15]. The second and third types serve as insulating and surveillance components of the nervous system.

1.2.2 Glioma Invasion and Difficulties in Treatment

Gliomas are brain neoplasms that originate from a single glial cell. Some brain tumours are benign, which means that they will not recur after treatment (surgical extraction and radiation). But most tumours are malignant and often continue to grow and diffuse even with treatment [96, 62, 26, 14], and ultimately lead to patient death. In general, malignant tumours outpace the growth of surrounding normal tissue as cancer cells divide and multiply more rapidly. A cell becomes cancerous after it undergoes some mutation (a chromosomal or genetic change) that results in an increase of the mitotic division in this cell or a decrease in a protein that suppresses cell division [55, 101, 27]. Eventually, cells that have undergone these genetic mutations become immune to inhibitory growth signals from their surrounding normal cells [101]. As mutations in malignant cells accumulate while these cells divide, tumours become even more invasive to normal tissue and threaten the patient's life. The speed and degree of invasion of malignant tumour growth are described in terms of tumour grades. There are four distinct glioma grades ranging from low-grade brain neopolasms (astrocytomas originating from an astrocyte glial cell) and intermediate (anaplastic astrocytoma) to the aggressive grade-four gliomas. These aggressive tumours, Glioblastoma Multiforme (GBM), are the most common among gliomas [30, 130]. The higher-grade malignant cells are more capable of invading normal tissue in a relatively short period of time [85]. The life expectancy of a patient that has been diagnosed with a GBM is usually less than two years in spite of advanced treatment methods [131]. Defining the tumour grade is essential for medical specialists as they decide the type and dose of treatment that will be delivered to the patient. Glioma treatment may combine one or more of three modalities: surgical extraction of the visible tumour, chemotherapy, and radiation therapy, the last often following surgery to kill the remaining border of cancer cells that have not been removed in surgery.

Despite substantial research efforts and more advanced treatment methods, tumours in general and high-grade gliomas in particular are still recurrent. Tumour recurrence may be due to several factors including the resistance of cancer tissue to treatment [28, 123, 111] or the presence of occult malignant cells that have not been removed through formal treatment [26, 77, 63, 133]. Hence, these occult cells start to invade adjacent healthy tissue, in particular around the tumour border though the visible tumour mass might have been removed or irradiated. The occult cells continue to diffuse in healthy brain regions beyond the tumour border, and form other small masses of cancer cells though the visible tumour may have been removed. Such small masses may remain undetected with MR imaging as they exist in small concentrations around the visible tumour. For example, even though surgical treatment eliminates the visible tumour mass, it is likely that the tumour will grow again from the diffuse cells and tendrils that have invaded the healthy tissue around the border of the original tumour [63, 133, 62] (*e.g.*, see Figure 1.3). These diffuse tendrils remain in low concentrations in normal tissue regions around the original tumour, and so are invisible on MRI scans but can be detected with spectroscopy. These low concentration tumour masses are usually eradicated with radiation therapy by irradiating a generous 2cm margin around the border of the original tumour [48, 56, 126].

In more aggressive tumours that grow into a significantly large mass, the malignant cells compete for nutrients, and so form a necrotic area at the core of the tumour as the cells at the centre are actually dead. However, the cancer cells at the periphery of the tumour border continue to proliferate adding to the slower-growing malignant mass, and usually secreting chemicals that make blood vessels grow into the tumour supplying it with nutrients (this phenomenon is known as angiogenesis) [40, 41, 42]. Malignant cells use nutrients and resources from healthy normal cells, exchange materials with normal cells, overload them with waste products, and are capable of compromising their survival in many ways [44].

Glioma treatment is usually effective in extending the patient's life by a number of months or years. But treatment, in particular with radiotherapy, can become more effective and may even grant the cure to many patients, if the treatment volume is specified in a way that would help eradicate the diffuse cancer cells and tendrils by applying a high radiation dose; therefore reducing the possibility of recurrence while minimizing the amount of healthy tissue compromised.

But specifying the treatment volume accurately is not a simple task. Despite advanced research, it is still very difficult and may often be impossible to track occult cancer cells and diffuse tendrils that have infiltrated through healthy tissue. This is because there are no clear boundaries between the cancer region and the normal tissue [48, 56, 126] which makes

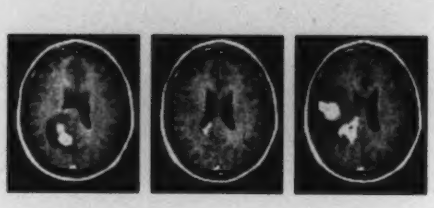

Figure 1.3: Example of glioma recurrence after treatment
This figure shows an example slice from each of three consecutive scans of a patient diagnosed with GBM, before and after surgical and radiation treatment. Each slice is obtained from a T1-weighted image volume with contrast, where the tumour shows as an enhancing mass surrounded by edema (swelling due to increased water content) appearing as a dark region. Left to right: the first slice (before treatment) shows the tumour in the posterior region of the right hemisphere adjacent to the ventricles. On the middle slice (after treatment), most of the tumour was removed and so, the edema decreased and is not visible on the image. The third slice, obtained from a later scan during patient follow-up, shows glioma recurrence in both the location of the original tumour and in regions adjacent to it. Tumour growth in adjacent regions is likely due to the infiltration of occult cancer cells through normal tissue as these cells build up into detectable masses over a period of time.

it unavoidable to irradiate healthy tissue along with the diffuse malignant cells resulting therefore in additional compromise of brain function.

1.3 Cancer research

Many health institutions have focused their efforts on cancer research leading to numerous advances in cancer treatment. But there seems to be no cure in the case of many patients; in particular those affected with higher-grade tumours. Because of this, cancer research has gone beyond the walls of medical institutions and biology labs to encompass research in mathematical and computational sciences as well.

Informatics and Mathematics Cancer research include but is not limited to automated tumour detection on MRI and CT scans, glioma growth modeling, statistical analyses of tumour growth rates, automatic tumour grading, evaluation of treatment efficacy, and treatment planning. Section 1.3.1 briefly mentions some of the recent work in cancer research, which motivates our approach that we overview in Section 1.3.2. These research models and several others are further described in Chapter 2.

1.3.1 Mathematical and Computational Modeling

Mathematical models are not new to cancer research. They have been proposed by mathematicians about half a century ago [109]. Early models were simple, and studied only solid tumour growth as the cell mass as a whole increased in diameter size. Some of these models assumed some sort of radial uniform growth. More recent and sophisticated models introduced the concept of *diffusion* as researchers started to develop better understanding of tumour growth and invasion. 'Diffusion' refers to the migration of malignant cells from the original tumour to healthy tissue adjacent to the tumour border. As these cells infiltrate throughout adjacent normal tissue, they build up into small tumours over time [85, 110]. Models of cell population growth were proposed by researchers in [125, 120, 121]. These models were based on age-dependence in cell proliferation and tried to predict the population size at a future time given the initial distribution for the cells. Research has also introduced logistic and exponential models [76, 108]. Logistic growth suggests that a portion of the tumour is proliferating while another portion is not, due to lack of nutrients as the tumour grows beyond nutrient availability. Early exponential models assumed that tumour cells proliferate exponentially and do not impose a restriction on the mass growth. These mathematical models and numerous others are further described in [85, 109].

Sophisticated computational models are possible now due to the many advances in algorithmics and computational sciences. In the last few years, many Informatics researchers have contributed to different areas of cancer research. Medical informatics contributions in the area of glioma research range from detecting the tumour on brain scans [95] to simulating its growth and modeling the treatment volume for radiotherapy [138]. Zizzari [138] proposes software modules for tumour modeling and radiotherapy planning. The tumour modeling module introduces a tool that predicts tumour growth, which is used for planning radiotherapy target prediction to help medical experts specify the treatment volume for conformal radiotherapy. Also, other researchers [21] tried to simulate tumour growth by introducing a diffusion model that captures the evolution of tumour density over time, and an expansion model that predicts the mass effect induced by tumour proliferation in the brain.

Other research projects have utilized data-driven approaches for decision making in cancer-related topics, other than tumour growth modeling, such as glioma grading [64], and robotic surgical planning in the brain [8]. But in this thesis, we focus on approaches to glioma diffusion modeling, which we compare to existing models (see Chapter 2).

1.3.2 Brief Overview of Proposed Approach

The purpose of our work is to examine the use of machine learning to allow more accurate predictions of future cancer diffusion in the brain of glioma patients. Our method is data-driven, and uses learning and decision-making to model glioma diffusion based on a number of features obtained from the data of several patients diagnosed with various glioma grades. We first process the patient data which consists of a sequence of MRI scans (see Chapter 3) and from which we extract the features for the learning model. We then learn a model that takes as input labeled brain volumes, described by features relevant to the patient and the tumour, and attributes local to regions adjacent to the tumour. We use the learned model to perform voxel-level classification (as described in Chapter 4.2) by assigning each unlabeled voxel around the tumour (at the first of two consecutive time points) to the 'tumour' or 'non-tumour' class given the features of the unlabeled voxel. We finally predict the tumour growth by running a diffusion algorithm that iterates sequentially over the voxels at the active border of the tumour using the classification labels to predict the new status (*i.e.*, at the second time point) of each voxel on the tumour-healthy border. This prediction is based on the probability of the voxel being classified as 'tumour' and on the number of *tumour*-voxel neighbours adjacent to the candidate voxel. The tumour border becomes enlarged as the algorithm iterates as more voxels become 'tumour'. The model is complete when the tumour reaches the volume at the second of the two consecutive time points. We evaluate our approach against the manually segmented *Gross Tumour Volumes* (GTV), and we compare the results to those produced by the standard uniform growth [48, 56, 126], and a tissue-based growth model that assumes faster diffusion in white matter [110]. We quantify performance for patient-specific predictions in terms of precision and recall [12].

1.4 MR Brain Imaging

Magnetic Resonance Imaging (MRI) provides a 3D description of internal brain structures by measuring at each pixel the following quantities: Proton Density (PD), spin-lattice (longitudinal) relaxation time (T1) and T1 with contrast (T1C), and spin-spin (transverse) relaxation time (T2), which are innate attributes of the tissue being scanned. To obtain contrasted images for the purposes of diagnosis and treatment, pulse sequence parameters, including T1 and T2, are varied. The contribution of PD to the image contrast depends on how many water molecules are present in the tissue being scanned (*e.g.*, fat, fluid, and edema are high PD regions). T1 and T2 depend on there being protons present to create a longitudinal and transverse magnetization. In this thesis, we mainly use T1-weighted and T2-weighted image volumes; see Figure 1.4 for an example slice of each of T1-weighted, T1 with contrast, and T2-weighted brain images.

A brain scan depends on factors including the pulse sequence, cost, time, resolution, slice thickness, inter-slice distance, noise signal, etc. The intensity values seen on an MRI scan for a particular brain depends primarily on the content of that pixel versus neighbouring tissue and on other factors including the presence of fluids (*i.e.*, cerebrospinal fluid) and fat. For example, on T1-weighted images, bone, air and water show as dark regions while fat has bright intensities. On T1-contrast scans, tumours show as bright regions because the contrast agent enhances the leaky blood vessels of tumours. On T2-weighted images, water and fluid have high signal while bone, air, and fat remain dark, which makes feasible the detection of both tumour and edema together on T2 scans [47, 57, 118].

More information about the physics of MRI and how the different parameters are measured could be found in [10, 47, 57, 118]. Coronal, sagittal, and axial views (see *e.g.*, Figure 1.4) can be acquired for each brain scan providing the full 3D volume of the brain. Each pixel in the image is referred to as a *voxel* speaking in terms of 3D volumes. Each voxel in a particular slice has a color intensity value that corresponds to a measurement of the tissue weighted by the MR parameters (PD, T1, or T2) averaged over a small 3D region [47, 57, 118]. Medical experts utilize their knowledge of the human brain anatomy and structure in conjunction with the intensity values visually detected on 3D MRI volumes to recognize abnormalities and tumours in the brain.

Data used in this thesis is limited to T1-weighted, T1-contrast, and T2-weighted MRI scans. Experiments have been performed only on the axial views of the MRI scans.

Brain imaging has a significant role in tumour detection and treatment [33]. Recent, more advanced imaging tools and methods allow radiologists and oncologists to examine the patient's brain on the 3D image volumes, to properly detect tumours, and to plan treatment based on detected glioma invasion. Imaging also allows doctors to track tumour

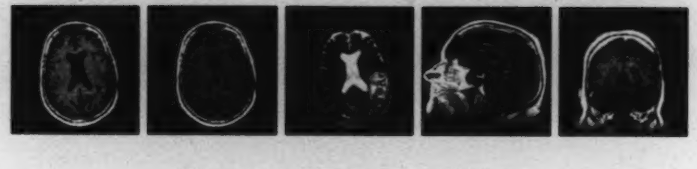

Figure 1.4: MR image modalities

Left to right: T1-weighted, T1 with contrast, and T2-weighted axial views, followed by the sagittal and coronal views of the same brain.

growth for follow-up purposes during treatment phases in order to examine the efficacy of the treatment method and to apply the treatment to regions where the glioma started to further diffuse. Patient follow-up is of particular importance during a radiotherapy treatment course where a sequence of brain images taken over a period of a few weeks or months helps medical experts measure tumour response to radiotherapy and track regions where the tumour has spread [33]. Doctors can then modify the initial treatment volume to include regions where more recent tumour invasion has been detected, and to apply to these regions an appropriate radiation dose as required. Section 1.5 illustrates in more detail radiation treatment for brain cancer patients.

1.5 Radiotherapy

Radiotherapy, surgery and chemotherapy, represent the three most significant and commonly-used treatments for malignant tumours including gliomas. In some cases, radiotherapy is used alone for tumour treatment but more often, it follows surgery or is combined with another type of treatment to kill the cancer cells. In this section, we only describe radiotherapy as the work in this thesis is motivated by its direct application to the treatment of gliomas using radiation therapy.

Radiotherapy is applied by delivering an appropriate dose to the cancer tissue while irradiating the patient from different directions with high energy photon beams generated from a linear accelerator. The different directions of the beams are generated in a way as to maximize the radiation dose at the targeted tumour volume while minimizing the dose delivered to the healthy tissue surrounding the tumour [94, 127, 128, 138]. In conventional radiotherapy, the treatment volume includes the visible tumour itself along with a 2cm margin around the visible tumour mass [48, 56, 126]. This 2cm margin may include both occult cancer cells and normal tissue alike, and may also spare occult tendrils that have already diffused beyond the treatment margin but remained undetected on the MRI scan due to low cell concentration. Tumour volumes of any arbitrary shapes are treated with radiation since the projected beams are customized into the different geometrical shapes of the treatment volumes. Customizing the beams allows medical doctors to vary the radiation dose administered to different regions of the brain as to maximize the dose delivered to the Clinical Target Volume (*i.e.*, the brain region clinically defined as diseased). Targeting the tumour with high radiation doses ensures a better chance for cure and a smaller chance for cancer recurrence [50, 138]. The total radiation dose is delivered in a treatment course through several fractions usually over a period of several weeks.

Though radiation therapy for gliomas proves to be effective because it can extend the patient's life, it has some very serious side effects due to radiation delivered to healthy cells, and in particular the region within the 2cm margin, causing reduction or even complete loss of brain function in the treated regions [19]. Radiotherapy may also impose the danger of initiating other malignant tumours in the irradiated organ (*i.e.*, in normal tissue adjacent to the treated region) as an irradiation side effect [135]. Radiotherapy is usually more effective in the first course of treatment than in a second or third course if applied to a recurrent tumour. But, due to radiation toxicity, it may not even be possible to treat the patient with an additional course of radiation therapy in case of glioma recurrence [32].

1.6 Machine Learning

In recent decades, machine learning algorithms have been widely used in performing a myriad of tasks including disease diagnosis, speech recognition, planning, robot control, prediction, etc. Traditionally-used machine learning algorithms range from simple, probabilistic classifiers like Naïve Bayes [36], which once learned from existing data, classifies a new data instance by calculating the posterior probabilities of a hypothesis, to more complex algorithms like Neural Networks that loosely simulate information processing by the human brain, and that involve feed-forward and error propagation algorithms [81].

In supervised learning tasks, we first learn a model from existing training data. Each data instance is often expressed as a feature vector that provides information with respect to this instance. We then use the learned model to perform a prediction by classifying new unlabeled instances such as each instance will be assigned a class label. The classification of an instance is associated with a decision function score or a probability. Results may be evaluated in terms of standard precision and recall [12].

In this thesis, the prediction task is about finding out whether a normal voxel, transitively adjacent to the border of the visible tumour, will become malignant. To perform this prediction task, we first learn a classifier from a large number of data instances obtained from several brain volumes of patients diagnosed with gliomas. Each data instance is represented by a feature vector specific to a voxel located in tissue adjacent to the tumour, and expresses information about the patient, the tumour, and attributes local to that voxel and to the neighbourhood of the voxel. After learning the model from patient data, the classifier will assign a label to each new unlabeled voxel, in regions adjacent to the tumour, that indicates whether that voxel is normal or diseased (*i.e.*, 'tumour' or 'non-tumour'). Each voxel is also assigned a probability as predicted by the classifier or a decision function score from which we estimate the probability that the voxel is 'tumour'.

The selection of an appropriate set of features is crucial to the learning and prediction tasks. The features are discussed in detail in Section 4.5. Other issues that emerge in learning and prediction are data bias, class imbalance, noisy data, selection of the learning dataset, data overfitting, etc, which are further discussed in Chapters 3 and 4.

1.7 Motivations for glioma growth modeling

The work presented in this thesis has direct application in radiotherapy for glioma treatment. As mentioned in the introductory paragraphs about glioma diffusion and treatment, the conventional method of treating glioma with conformal radiotherapy is to irradiate a 2cm margin around the visible tumour mass, which typically includes both healthy tissue and malignant tendrils. This radial 2cm margin does not take into account any information about the patient or the tumour but simply assumes that occult malignant cells have uniformly diffused into surrounding healthy tissue. Because this conventional treatment method is generically applied to all patients regardless of where the cancer has actually diffused, it often compromises brain function and may still leave behind cancer tendrils allowing tumour recurrence. Hence, the necessity for modeling glioma diffusion – assuming that occult cancer cells that have infiltrated through adjacent healthy tissue are responsible for further tumour growth – and finding new methods that may help radiation oncologists specify more accurately the appropriate treatment volume and radiation dose. In addition, the introduction of an accurate glioma diffusion model provides the advantages of increasing the level of precision and standardization in specifying the treatment volume, which will also help oncolgists save time and effort required for treatment planning.

This research investigates how machine learning algorithms can be applied to predict tumour growth by modeling glioma diffusion in order to ultimately help identify the border of treatment volumes in glioma patients. A more accurate specification of the treatment volume can help improve conformal radiotherapy by minimizing the chance of tumour recurrence, and sparing healthy tissue around the tumour thus saving brain function. The diffusion model proposed in this thesis is based on the labeling of a probabilistic classifier that learns a model from a set of features specific to the patient and the tumour, and attributes of the diseased region around the tumour.

1.8 Scope of Thesis

This thesis presents a study of glioma growth modeling. The main results consist of a general learning model that may provide insight into glioma growth dynamics and that would supply medical experts with a tool to help them specify the treatment volume for irradiation. The application of the model is not restricted to a particular tumour type, *i.e.*, the model can be applied to all brain tumours ranging from grade-one astrocytomas to the notoriously aggressive glioblastomas (GBM). The proposed diffusion model involves using a classifier to predict whether a voxel adjacent to the tumour will become cancer. Our classification is based on a number of features specific to the patient and the tumour and attributes of the voxels that are recursively adjacent to the visible tumour. A diffusion algorithm iterates around the active tumour border at an initial time point, and assigns eligible voxels (*i.e.*, that are transitively adjacent to the tumour border) to the 'tumour' class using the labels predicted by the classifier. The algorithm terminates when it has added to the initial tumour volume a number of voxels that represents the set difference between the initial and final tumour volumes. The output of the diffusion model is a contiguous 3D tumour volume that corresponds to the GTV at a later time.

However, the actual specification of the treatment volume and the radiation dose is outside the scope of this research study as these pertain to the medical side of this work, and are the responsibility of medical specialists.

This dissertation confirms the thesis that our CDM system performs more accurately in most cases than radial uniform diffusion, UG, and tumour growth based on the heterogeneity of brain tissue, GW. To demonstrate system performance, we compare results numerically and analytically (see Chapter 5).

1.9 Thesis Outline

This first chapter has provided an introduction to cancer in general and gliomas in particular. It also described current glioma treatment methods, with a section on radiotherapy, and described the advanced tools and methods in cancer research to help improve treatment, in particular radiation therapy. This chapter also includes a section on machine learning and how the learning and classification tasks are utilized in glioma diffusion modeling.

Chapter 2 presents an overview of glioma treatment and the challenges involved in the treatment task. It also describes previous and current work in brain cancer research with emphasis on mathematical and computational models, and their usefulness in glioma treatment and in improving the effectiveness of radiation therapy.

In Chapter 3, we describe the model framework including the data processing pipeline, the feature extraction, and the classification and modeling tasks.

Chapter 4 covers the prediction task including the system implementation details and the use of machine learning algorithms. This chapter also describes the feature space as well as the training and testing data folds.

Chapter 5 presents the results and compares the output of our CDM model, evaluated against manually segmented *GTV*, with the output of the uniform radial growth used in conventional radiotherapy and of the tissue-based growth model that assumes a 5 : 1 diffusion ratio in white versus grey matter.

Chapter 6 concludes the thesis and presents ideas for future work and improvements to our current system.

Chapter 2

Related Work

In this chapter, we describe conventional treatment methods of gliomas as well as mathematical models that have been implemented to predict glioma growth. Section 2.1 describes clinical issues and challenges involved in the treatment of glioma patients. Section 2.2 describes existing mathematical and computational approaches that were developed to illustrate tumour growth behaviour and to predict glioma diffusion. Section 2.3 discusses and evaluates standard glioma growth models, which motivates the implementation of our proposed approach. The chapter concludes with a summary of issues and concepts discussed in these sections.

2.1 Glioma Treatment

Glioma treatment has represented one of the most challenging tasks in cancer research in the past few decades tough gliomas do not usually metastasize outside the brain (likely due to the blood-brain barrier). The efforts of thousands of researchers have been dedicated to understanding and explaining brain neoplasms, and to discovering new methods of treatment or at least improve conventional methods. Both theoretical and experimental research has shed some light on glioma growth and treatment, and has helped identify some of the factors involved in cancer diffusion and that make this disease difficult to cure. Not only are gliomas diffuse and invasive making efficient treatment a challenge, but also some of the factors involved in the treatment of these neoplasms include the biological and genetic complexity of the tumour, the growth rate, and the grade. It is known that some tumours may have histologically heterogeneous cells capable of resisting treatment [85]. In addition, high-grade malignant cells proliferate much faster, and are more invasive to neighbouring normal tissue. Such cells deprive healthy cells from nutrients, overload them with waste components, and compromise their survival. Unfortunately, these high-grade gliomas, known as glioblastomas (GBM), are the most common among brain tumours (see [30, 44, 85, 109, 130]).

There are several treatments for gliomas, which mainly include chemotherapy, radiation therapy, and surgical resection. Surgery has been used over the centuries to extract tumour masses [85]. The resection of the tumour through surgery provides the benefit of relieving the pressure in the patient's brain and reducing the symptoms induced by the presence of the tumour. Surgery is effective when removing benign tumours since such tumours usually do not recur. But in the case of malignant tumours, surgery alone is ineffective. This is because it is impossible for the surgeon to remove the malignant tumour without leaving behind a border of cancer cells [96, 62, 26, 14, 85]. This is why other treatment methods are often used after the surgical operation. These treatment methods include chemotherapy and radiation therapy. Chemotherapeutic treatments are drugs triggered by the cell cycle phases. These drugs use specialized chemicals to eradicate the malignant cells while the rest of the brain remains defended through the blood-brain barrier. However, malignant cells may develop drug resistance in particular because of cancer cell heterogeneity, which makes chemotherapy much less effective [85, 111, 113]. Radiation therapy has been conventionally used over the last few decades, and often follows surgery to kill the remaining border of cancer cells left after surgical resection.

Radiotherapy is a specialized science that has progressed quite significantly in recent years as experts have improved conventional methods and introduced new ones to properly target the malignant mass while minimizing damage to healthy tissue. Some of these new methods include computer-assisted radiotherapy in which algorithms and software modules are used to specify the treatment volume for irradiation [138]. Other software tools assist in viewing the brain in 3D giving a more visible, and detailed image of the diseased region, and help medical staff in treatment planning [22, 65].

In radiotherapy, the patient receives the appropriate radiation dose through beams coming from different directions in a way that concentrates the radiation dose delivered to the tumour and that minimizes the dose delivered to surrounding healthy tissue [138]. Radiation therapy is effective but has many side effects including brain necrosis [19], toxicity, and initiation of other malignant tumours in healthy tissue [135]. Radiation is usually effective in the first course of treatment. But it is not appropriate to reapply radiotherapy to a recurrent tumour after a first treatment course because of radiation toxicity, as healthy brain tissue surrounding the tumour may not tolerate another radiation dose [16, 18, 25, 32]. However, it is clinically known that a higher radiation dose delivered to an accurately specified treatment volume will ensure a better chance for cure and will reduce the chance of malignant recurrence [50, 138]. It is also known that tumours are more likely to recur when there is a high density of undetected cancer cells in the surrounding normal tissue or after surgical resection due to the thousands of cancer cells left around the periphery of the already extracted tumour mass [96, 62, 26, 14]. In this way, tumours have been compared in an analogous manner to a forest fire where the action of further spreading is happening at the periphery rather than at the core of the fire [85].

Though these main treatment methods have helped extend the lives of thousands of patients and perhaps provide the cure for some of these patients, current methods have often fallen short of properly targeting the action of tumour growth and invasion, which leads to malignant recurrence and further diffusion; the latter being even more difficult to treat.

Mathematical and computational modeling research has offered significant contributions to cancer research in general over the past few decades and has helped provide better understanding of tumour diffusion dynamics. Many of these mathematical models are overviewed and cited in Swanson's thesis [109]. Several of these models have been implemented to predict or model tumour growth, to plan treatment and to help specify the target volume for radiotherapy. Other models were introduced to help quantify the effect of treatment on tumour growth, to determine optimal treatment methods, and to study the effects of various therapy strategies including combined treatments on specific tumours [111, 113]. Computational models with a learning component, in particular models that involve data-driven learning, decision making and a prediction component, are the most recent and offer much promise in understanding tumour growth, and in improving current treatment modalities (see *e.g.*, computer-assisted radiotherapy [138]).

The rest of this chapter presents an overview of mathematical models of brain tumour growth, statistical analyses of glioma diffusion, and computer-assisted treatment plannning. The following section of this chapter covers in detail each of these models and illustrates how the standard approaches relate and compare to our proposed model.

2.2 Tumour Modeling

Current mathematical and computational models of interest can be grouped into two main categories: *microscopic* models at the sub-cellular level describing internal activities within the cell and at the cellular level speaking of cells division, proliferation and interaction, and *macroscopic* modeling which studies overall tumour propagation, diffusion and the increase in cell density.

Earlier models were simple and merely based on exponential tumour growth (see an overview of these models in [85, 109]). These models were later modified to account for the gradual slow down of the tumour growth rate when the neoplasm grows beyond a certain size or when the tumour mass becomes significantly large to form a necrotic centre [85, 109]. Other models – in an attempt to provide a better explanation of tumour growth behaviours – used cellular automata [59] to represent a more realistic diffusion process.

In this thesis, we model glioma growth at the voxel level where each voxel represents 8mm^3 on the image volume. Our growth model takes into account attributes of the patient and the tumour, and features of the voxel and of regions adjacent to the voxel. We initialize our model from the GTV at an initial time point, and we predict (using learning and classification) the *macroscopic* volume of the tumour that corresponds to the GTV at a later time point.

2.2.1 Macroscopic and Volumetric Modeling

Mathematical modeling of gliomas at the macroscopic level has represented the traditional framework in predicting glioma growth and diffusion, and has been conventionally recognized as a practical method of quantifying and predicting volumetric tumour growth in order to help improve treatment techniques. Macroscopic modeling is specifically useful in identifying the tumour boundaries at a future time by using growth and proliferation parameters.

We describe three standard glioma growth models at the macroscopic scale. The first model uses cellular automata to model the different states (or changes) of the tumour cells [59]. The second is a more recent model that presents three hyperbolastic growth models that extend early mathematical growth functions [116]. The third model, also relatively recent, is based on a diffusion-reaction equation, and was the first to include a supervised learning component for predicting the treatment volume [138].

Simulating Glioma Growth with 3D Cellular Automaton

The growth model presented by Kansal et al. [59] describes macroscopic tumour behaviour by simulating the *gompertzian growth* [59, 11]. The 'gompertzian model' views the tumour as a population of cells and the growth as a dynamic process in which proliferating and inactive classes of cells interact. The *gompertzian growth* can be typically modeled by non-linear differential equations describing the interactions of proliferating versus inactive tumour cells. In this model, Kansal et al. use time-dependent parameters and constant parameters over three orders of magnitude in radius. These parameters include the probability of division, the average tumour radius, the proliferating rim thickness, and the necrotic base thickness. Changes within the tumour mass are described by simulating the fraction of dividing cells, the fraction of cells that become non-proliferating, and the necrotic regions. This allows representing the transition of tumour cells from active, dividing cells (at the periphery of the tumour) to the necrotic state (in the centre of the tumour mass). The simulation is designed to predict the rate of growth as the volumetric doubling time at given radii, using an underlying lattice with variable grid size to allow the modeling of tumour growth over different orders of magnitude. The results of the simulation are obtained by averaging the radii of the cells at the edge of the tumour and of the necrotic region. The model determines the probability of division of active cells and the space radius with which cells divide or become non-proliferating. The volume and radius of the tumour are plotted as a function of time. The type of cell (dividing, non-proliferative, or necrotic) in each grid of the lattice is saved at the given times. The model acknowledges that tumours do not grow into perfectly spherical volumes, and allows the radii to vary over the surface of the tumour.

In comparison with earlier, more simplistic models, Kansal et al.'s model [59] provides a more realistic illustration of tumour growth behaviour in 3D at the macroscopic scale, and also suggests useful simulation parameters that take into consideration the different states of the cells inside the tumour. The model is evaluated, however, with one test case that supports its validity. Also, the model is designed to simulate homogeneous tumour pathology, *i.e.*, it does not take into account the tumour grade and the genetic variety of tumour cells. Therefore, the generality of the model remains in question – *i.e.*, the simulation [59] may not be able to represent low-grade tumours, and tumours with heterogeneous cancer cells – but it is acknowledged that the *gompertzian growth* is observed in GBM [138]. While Kansal et al.'s model does not assume radial uniform growth (as clearly stated by the authors), the simulation may likely produce spherical tumour structures because it is mainly concerned with active versus non-dividing cells within the tumour. The model is not designed to describe the action of malignant diffusion and infiltration of cancer tendrils through surrounding normal tissue. In this, it implicitly assumes that the tumour is an encapsulated mass with radial growth. Also, the simulation parameters do not express at any level the heterogeneous brain anatomy. In reality, however, cancer tumours grow into non-uniform masses and the malignant cells diffuse beyond the original tumour mass.

As opposed to Kansal et al.'s model [59] that describes the tumour in terms of proliferating, inactive and necrotic cell classes, our proposed model, CDM, predicts glioma diffusion in regions adjacent to the active border of the tumour. CDM assumes that occult cancer cells that may have infiltrated through adjacent normal tissue are responsible for further tumour growth. CDM does not use already-computed parameters or radii (as in [59]), but it 'learns' glioma diffusion patterns from patient data based on attributes describing the tumour and the tissue adjacent to the tumour.

Hyperbolastic Tumour Modeling

The methods described in [116] represent some of the most recent mathematical models in macroscopic tumour growth research. The three proposed growth models in [116], called 'hyperbolastic models' are designed to help predict and analyze volumetric, self-limited growth behaviours of multicellular tumours. These hyperbolastic models extend earlier methods [46, 92, 124, 129] that have been applied to several biological and medical studies. The proposed models [116] do not assume radial uniform growth, simulate asymmetric growth patterns as in real tumours, and also accommodate the concept of increasing versus decreasing tumours. However, these methods do not account for the various factors involved in clinical tumour growth behaviour such as the infiltration of malignant cells throughout normal surrounding tissue. Instead, Tabatabai et al.'s models [116] assume that tumours are encapsulated, self-limited systems, not incorporating therefore the interaction between the tumour and the adjacent normal tissue as well as the competition of cancer cells within the malignant mass. Because of these simplistic assumptions, these models may not realistically represent clinical tumour diffusion behaviour. In addition, these models are general to all types of tumours, *i.e.*, they are not specifically designed to model gliomas, and therefore they do not take into consideration the various structures of the brain anatomy.

Macroscopic Modeling and Supervised Treatment Planning

In his thesis [138], Zizzari contributed to glioma research with modules for three different research areas: tumour detection, tumour growth modeling, and treatment planning. His study of tumour growth modeling focuses on the development of a mathematical model that describes tumour growth dynamics on the macroscopic scale. His approach integrates a reaction-diffusion model, and uses parametric surfaces, compromising between the accuracy of the growth description and the computational complexity of the involved methods. In this work [138], Zizzari describes tumour proliferation using tensor product splines and differential equations, the solutions of which give the distribution of tumour cells with respect to their spatio-temporal coordinates. The model has been designed and implemented specifically to consider only the relatively simple volumes of glioblastomas (GBM). Zizzari suggests that his proposed method can be extended to be generally applied to the modeling of other evolving objects [138] but his model has not been applied to lower-grade tumours.

In addition to the macroscopic modeling of GBM tumours, Zizzari also introduced a third module for automatically planning the treatment volume [138]. The purpose of introducing this additional module is to provide a supporting tool that defines the treatment volume for conformal radiotherapy in a more precise and standard fashion as compared to conventional methods. This treatment planning module combines the use of both the reaction-diffusion equation and a supervised learning task with a feed-forward neural net-

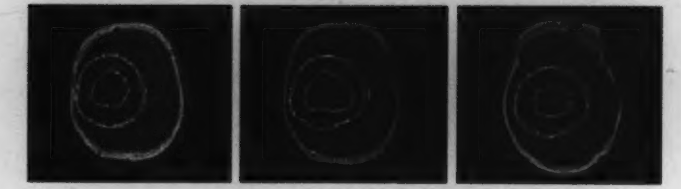

Figure 2.1: Zizzari's prediction of PTV [138]

Experimental results for Zizzari's GBM growth model on CT brain images: yellow lines are the GTV and PTV, respectively, obtained through a medical expert; adjacent white lines are the predicted PTV from Zizzari's proposed model. (Figure obtained from [138].)

work [138]. That is, the artificial neural network describes an appropriate radiation volume according to the detected areas on radiographic CT images, where the GTV has been manually segmented, and according to the expected growth dynamics described by the reaction-diffusion equation. To illustrate, the reaction-diffusion equation models the future expected tumour growth, leading to an estimation of the *Clinical Target Volume* (CTV). Then, to predict the *Planning Target Volume* (PTV), the neural network is applied to the same CTV such that the control points of the tensor product spline that approximates the surface of the computed CTV serve as input information to the neural network. The output of the neural network is a set of control points for the PTV that are associated with another tensor product spline, which describes another surface to approximate the predicted PTV (see Figure 2.1). The association between the input control points for the CTV and the output control points for the PTV has been established through *training* the neural network on samples extracted from the CT images, manually segmented by human experts (as described in [138]). The automatically predicted PTV has been compared to the PTV defined by a medical expert (see Figure 2.1).

Zizzari's work [138] provides the advantage of predicting the treatment volume in addition to modeling the tumour growth. But his model, like most traditional macroscopic tumour modeling approaches, attempts to quantify future tumour growth from volumetric and geometrical perspectives without considering biological factors – patient information, brain anatomy, tissue heterogeneity, pathology, etc. – that have implications in tumour diffusion. This is because his model was designed to specifically predict GBM diffusion only. Therefore, though the proposed approach represents reasonable descriptions of macroscopic tumour diffusion, it lacks a general method that could be applied to all tumour types, and it is still too simple to explain the clinical interaction of the tumour with its surrounding environment.

2.2.2 Glioma Modeling Based on White Matter Invasion

Unlike the traditional macroscopic and volumetric models that are mainly concerned with describing the tumour volume as a self-encapsulated system, the trend in glioma research is to study the factors involved in glioma diffusion. Many recent models incorporate anatomical features to help model tumour growth with a more realistic view based on the characteristics of the heterogeneous brain tissue. Some of these models extend the conventional modeling of gliomas at the macroscopic scale to explain tumour invasion at the microscopic, cellular level.

In this section, we describe glioma diffusion models that take into consideration the heterogeneity of brain tissue and models that use an anatomical atlas to recognize areas of malignant invasion. These models show that gliomas are more invasive in white matter tissue and that malignant cells may likely follow white fibres, which serve as a 'highway' to the proliferation of cancer cells while the invasive action of the tumour cells tend to be slower in grey matter.

Differential Motility in Grey and White Matter

The model presented by Swanson et al. in [110] models glioma diffusion based on the differential motility of gliomas in grey matter versus white matter. Differential motility is estimated using Fisher's approximation [84, 14] and is based on the suggested linear velocity of glioma cells in [17] from which are estimated the diffusion rates of glioma cells in grey matter as compared to white matter. Swanson et al. suggest in [110] that microscopic tumour invasion is not detectable on MRI scans, and that it is therefore important to identify pathways of glioma invasion in order to help model tumour growth more accurately, which would then have a direct implication in specifying more precisely the treatment volume for irradiation. Swanson et al.'s model [110] extends a basic mathematical model of glioma growth and diffusion based on the analysis of serial CT scans taken in the terminal year of a patient with recurrent anaplastic astrocytoma (see [29, 119, 133, 14]). The following equation summarizes this basic mathematical model:

Rate of change of tumour cell density =
diffusion of tumour cells + growth of tumour cells (⋆)

Swanson et al. extend this basic equation to account for the heterogeneity of brain tissue in their new model [110]. In this equation, the diffusion coefficient D will have two different values – *i.e.*, the diffusion coefficient D is a function of a spatial variable differentiating regions of grey and white matter. This function will express the faster malignant invasion of tumour cells in the white matter as compared to diffusion in grey matter, *i.e.*, $D_w = 5D_g$. The diffusion parameter is estimated from experimentally observed linear velocities in grey and white matter (see [17, 84, 14, 110]). The five-fold difference in the diffusion coefficients in grey and white matter is used throughout the simulations [110]. More details about the computations of equation parameters and the framework of the model are found in [109, 110, 85]. Swanson et al.'s model offers an important contribution by accounting for the heterogeneity of brain tissue. But it is not clear how the model has been evaluated as there are not proper comparisons with clinical data.

Swanson et al. extended the work in [110] to simulate virtual gliomas as described in [112] with the purpose of evaluating the effectiveness of medical imaging and current glioma treatment. This virtual glioma model [112] quantifies the spatio-temporal growth and invasion in three-dimensional space on a virtual human brain taking into account the gross anatomic boundaries available through the BrainWeb [24]. Swanson et al.'s model [112] is described as the rate of change of tumour cell population density – see the above equation ($\star$) in this section. The simulation allows the virtual tumour to grow and migrate within a 3D representation of the brain. The proposed model also tracks progression of tumour growth from a diagnostic size of 3cm to a fatal size of 6cm in diameter. The model is relatively simple as it focuses mainly on diffusion and proliferation excluding other factors involved in tumour growth. The model is also easy-to-understand, and provides the flexibility of varying different parameters though there is no quantitative measure of evaluating the truthfulness and accuracy of the simulation.

Sawnson et al. used their original model in [110] to measure the effectiveness of chemotherapy in both cases where chemotherapeutic drugs are delivered homogeneously to the brain and in heterogeneous quantities to different brain tissue types [111]. This new model is based on identifying white matter as favorable tissue for tumour growth and invasion while grey matter is not [110]. The model explains that treatment is most effective in certain brain regions such as the cortex and is least effective as the tumour moves radially inward from the cortex and as white matter length increases. The model also illustrates tumour growth over time both with and without chemotherapy, and shows that even for fairly effective chemotherapies, the tumour continues to grow though this growth is slowed down when chemotherapeutic drugs are being delivered to the patient. White matter invasion is clinically undetectable at first due to low concentration of the cancer cells but after a period of time, cancer cells build up into a detectable mass. This model is particularly important as it emphasizes the significance of taking into account the heterogeneity of brain tissue when modeling tumour growth behaviour, treatment response, and recurrence.

In addition, Swanson et al. made use of their mathematical modeling of gliomas [110, 111, 112, 114] to illustrate the limitations of current imaging techniques, to explain recurrence, and to suggest therapeutic intervention in [115]. Swanson et al. review several of their findings in [113] where they also describe the modeling of patient survival time, the modeling of lower-grade gliomas, and the modeling of chemotherapy and surgical resection. They also discuss the mapping and interpretation of their simulation results to the threshold of detection on CT and MRI brain images (see detailed descriptions of these various models and model framework in Swanson's thesis [109] and in Chapter 11 of [85]).

White Matter Tract Invasion

Diffusion-Weighted Imaging (DWI) is a magnetic resonance technique that is sensitive to the movement of water molecules [75]. Diffusion Tensor Imaging (DTI) is a modification of DWI that is able to depict the anatomy of the white matter fibre tracts in the brain.

Price et al. [90] use T2-weighted scans and DTI sequences for 20 patients to determine whether DTI of brain tumours can demonstrate abnormalities on T2 scans. Regions of interest were drawn within the tumour as well as in white matter at distance from the tumour and in areas of abnormality on DTI that appeared normal on T2 images. Results show that abnormality detected on DTI was larger than that on T2 in most patients with high-grade gliomas [90]. Price et al. conclude that white matter disruption can be identified using DTI for high-grade glioma patients, changes in DTI may be due to tumour infiltration, and that DTI may provide a useful method for detecting occult white matter invasion.

Clatz et al. [21] propose a model that simulates the growth of Glioblastoma Multiforme (GBM), the most aggressive and common gliomas. Clatz et al.'s model assumes that gliomas grow faster in white fibres in the brain. Their model is based on an anatomical atlas that includes white fibre diffusion tensor information and delineation of brain structures. Clatz et al. model tumour invasion with a diffusion-reaction equation and the mass effect induced by the tumour growth with a linear elastic constitutive equation. Model results are based on two consecutive axial scans acquired within a six-month time interval for the same patient. The first scan is used in initiating the simulation and the second in measuring the model's accuracy. The model offers the advantages of patient-specific simulations and of using diffusion tensor information in tracking white matter invasion.

We currently do not have access to DTI data that we could incorporate into our CDM system, but as suggested by Clatz et al.'s model [21], using DTI data represents a promising research direction as it helps more accurately indicate cancer invasion in white matter fibres.

2.2.3 Statistical Analyses of Glioma Diffusion

In this section, we describe some of the existing statistical studies that were performed to analyze tumour diffusion and to quantify the capability of malignant cell invasion. The first study measures the continuous growth of the mean tumour diameter in untreated grade II gliomas using linear regression [73]. The second study is a test of two algorithms, nearest-neighbour and surface modeling, to quantify the tumour volume and measure the growth rates [49]. We also briefly mention the use of statistical modeling in measuring tumour mass effect on anatomical deformations in the brain [83].

Estimating the Tumour Growth Curve with Linear Regression

Mandonnet et al. performed a statistical study [73] in which they reviewed MRI scans of 27 patients with untreated grade II gliomas. In this work, they present a model for proliferation and diffusion of malignant cells to predict the linear growth of the mean tumour diameter. Mandonnet et al. performed linear regression statistical analysis to estimate the average slope of the growth curve of the mean diameter for low-grade oligodendrogliomas and mixed gliomas. Results show that the mean diameter of the tumour varies among patients and grows linearly during the follow-up period, at a predictable rate in its pre-malignant phase, but does not obey exponential evolution. The reported average slope of the evolution of the tumour mean diameter is $0.00113cm/day$; this diameter increases by $4.1mm/year$. Mandonnet et al.'s statistical analysis has the advantage of access to the data of *untreated* gliomas, which is not a usual opportunity. Because the data in this study is not affected by treatment (*i.e.*, as opposed to treatment data often including tumours decreasing in volume), parameters in Mandonnet et al.'s model can be reliably used in glioma diffusion and radiological growth modeling.

In our model, however, we did not use the growth rate suggested by Mandonnet et al. for two reasons. First, this growth rate is relevant to only low-grade and mixed gliomas (*i.e.*, Mandonnet et al.'s study does not include GBM) while our data represents various tumour grades including high-grade gliomas. Second, the above growth rate is measured for *untreated* gliomas while our data consists of studies of *treated* glioma patients where tumour growth rates vary depending on patient response to treatment.

Tracking Growth Rates with Nearest-neighbour & Surface Modeling

In their study of GBM growth [49], Haney et al. report the results of applying two 3D image analysis algorithms to serial MRI scans. The algorithms, nearest-neighbour tissue segmentation and surface modeling, were applied separately to the scans to track the rates of volumetric change for contrast-enhancing tumours. The outputs of the algorithms were compared to a manually defined standard of reference. Haney et al. claim, according to their analysis results, that growth rates obtained with nearest-neighbour algorithm – measured in terms of halving and doubling times – were highly correlated with the standard of reference, while those obtained with surface modeling were not, even though surface modeling algorithm is reported to reliably quantify tumour volumes. Growth rates are calculated from a time factor – doubling or halving – and from the interval time in days, and the volumes at onset and the end of the interval period respectively. The significance of this work is in part the segmentation and quantification of the changing tumour volume on the MRI scans, and tracking glioma growth rates, which helps determine patient response to treatment. Also, model parameters could be used in extending supervised pattern recognition approaches to predicting potential tumour diffusion based on data in the MRI scans (Chapter 4).

Modeling Mass Effect based on Biomechanical Simulations

The work in [83] describes the effect of the increasing volume of the tumour mass on deforming the surrounding brain tissue. This effect has been modeled by employing biomechanical simulations of tumour growth that use the Abaqus CAE [1] environment. These tumour growth simulations served in initiating a statistical model for predicting the anatomical deformations induced by the tumour growth in the brain [83].

2.3 Discussion

As described in Section 2.2, the glioma diffusion models presented so far in the state-of-the-art literature have been designed to mainly describe the geometrical growth of gliomas as evolving objects. Many of these models introduced macroscopic descriptions of tumour diffusion behaviour in terms of cell exponential growth, dividing versus necrotic cells, and the proliferation of the malignant mass as a whole. Some of these models have modeled tumours as self-limited, encapsulated systems, not taking into account the biological complexity of the tumour cells, the interaction of cancer cells with surrounding normal tissue, and the heterogeneity of the brain anatomical structures (see Section 2.2.1). These models are not capable of tracking the diffusive, invasive action of malignant cells as they infiltrate through adjacent healthy tissue. Also, these models tend to yield symmetrical or spherical structures of the predicted tumour growth.

Other models incorporated anatomical factors that are believed to have a role in glioma diffusion. These factors include the different motility rates of cancer cells in heterogeneous brain tissue [110], and the more invasive nature of glioma cells as they spread through white matter tracts [21, 90]. Anatomical factors considered also include the different brain structures that represent pathways of faster diffusion versus boundaries where tumours cannot spread. These models are likely to be more reliable than the macroscopic models described in Section 2.2.1, and they will more closely predict the asymmetric structures of tumour growth as clinically observed.

Statistical studies helped provide some insights into the understanding of glioma diffusion. These statistical analyses (Section 2.2.3) were usually carried out to quantify glioma diffusion and to track tumour growth rates in both cases of untreated [73] and treated [49] gliomas.

It is worth noting that all of these models have been mainly concerned with specifically predicting the diffusion of GBM, the most invasive type of gliomas. Then, researchers have tried to extend their high-grade glioma predictions to model low-grade or mixed gliomas [85]. Consequently, the literature lacks a general model that is capable of predicting glioma growth behaviour without restriction to the tumour grade or pathology. In addition, all models presented so far are evaluated by visual comparisons of the clinical tumour growth or by numerically comparing model parameter values with experimentally reported values. Neither of the researchers has performed thorough evaluations of glioma growth prediction in terms of precision and recall measures [12].

Based on the above observations and the evaluations of the existing approaches, we conclude that simplistic models are not capable of predicting clinical tumour diffusion and it is therefore essential to identify tumour properties and biological factors that help explain glioma diffusion. Based on our observations from the state-of-the-art literature and from contemporary medical knowledge, we identify several factors that will contribute to more successful glioma diffusion modeling:

- The tumour grade – might be implied by the growth rate since high-grade tumours tend to grow much faster than lower-grade tumours as clinically observed
- The brain tissue – grey versus white matter
- The brain anatomy – regions that represent pathways versus brain structures that act as a boundary to the spreading action of the malignant cells
- The neighbourhood at the periphery of the tumour where there is always interaction between the malignant and the normal cells. These interaction regions where cancer cells have likely diffused are characterized by the presence of edema and appear as enhancing regions on T2-weighted images.
- The location of the tumour inside the brain – decides the shape of the tumour depending on surrounding anatomical structures, and helps estimate growth regions where the tumour has induced anatomical deformations due to mass effect.

We incorporate these diffusion factors as learning features into a 'general' diffusion model, CDM, that we present in this thesis. Our CDM model is a learning system that incorporates the above factors represented in terms of learning features and also includes several other image-based and tumour-specific features that account for the spatio-temporal aspect of the diffusion process (see implementation details in Chapter 4). Unlike some existing approaches that implicitly assume radial or symmetrical growth and try to predict the macroscopic geometrical evolution of the tumour, our model does not make any assumptions about the shape or growth pattern behaviours of the tumour. Neither does it insist a-priori on the diffusion rates in grey versus white matter as in [110]. Instead, CDM learns the glioma diffusion behaviour based on the training patient data and the learning features as opposed to simulating virtual tumours. This learning task acts on the voxel level by training on a large number of voxels – at the tumour periphery – obtained for several patients. This makes our prediction model more accurate as we consider voxel-specific properties in the training fold and voxel-specific labeling in the classification component, rather than a rough modeling of the tumour as an evolving object, conventionally implemented in macroscopic prediction approaches. It is worth emphasizing that unlike Zizzari's learning method [138], CDM does not learn control points in a predicted approximation of the treatment volume, but it learns the factors and features that contribute to tumour diffusion. Therefore, CDM is a more sophisticated learning system that describes more tangibly the characteristics of the tumour and of the diseased region tracking potential diffusion according to features specific to the tumour and to the vulnerable region adjacent to the tumour.

We perform patient-fold testing and we quantify the accuracy of our predictions in terms of the standard precision and recall measures, which are more accurate than visual and graphical comparisons of model results against clinical tumour growth as done in [21, 138]. We initialize the model from the manually segmented GTV on the MRI patient scan at one time and we evaluate model performance from the segmented GTV of the same patient at a later time (within a few months interval).

CDM is a general model that can be applied to tumours of different grades, as opposed to GBM-specific modeling in existing approaches. But the model does not offer treatment planning and radiation dose specifications as these are left to the discretion of medical experts.

2.4 Summary

In summary, gliomas are usually treated with either of chemotherapy, radiation therapy, surgical resection, or a combination of these treatment methods. Unfortunately, conventional glioma treatment methods have fallen short of providing the cure to the large numbers of patients; in particular those diagnosed with high-grade gliomas. This is why thousands of researchers around the world have dedicated their efforts to studying and modeling gliomas in order to improve conventional treatment modalities or to discover new treatments.

Numerous mathematical models have been implemented in an attempt to understand the factors that play a role in tumour growth. Some of these models were simple and made assumptions with respect to tumour growth behaviour such as radial growth patterns and self-limited, encapsulated tumours. These methods modeled the tumour at the macroscopic scale as an evolving object but did not account for the biological complexity within the tumour cells nor for the heterogeneity of brain tissue. More recent mathematical models are more sophisticated and take into consideration the biological factors involved in tumour growth and invasion.

In Medical informatics, research areas range from the development of tools and algorithms that are capable of automatically detecting the tumour volumes (GTV) on the MRI or CT scans to predicting cancerous tumour diffusion. Research studies also include but are not restricted to disease diagnosis, the quantification of drug efficacy, the development of treatment planning modules, and the improvement of conventional treatment methods.

As observed through the sections of this chapter, there exists many methods that use mathematical approaches to quantify or predict tumour growth. Some methods are limited to modeling virtual tumours rather than using clinical data, which might be due to the difficulty of obtaining patient data. Other methods that use clinical data, whether to initialize a simulation or to measure model accuracy, are often implemented to specifically model glioblastomas (GBM), the most common and aggressive brain tumours. So far, the state-of-the-art literature did not present a general glioma growth model that would be able to predict malignant diffusion in different tumour pathologies without limitation to a particular tumour grade.

To the best of our knowledge, our CDM system is the first *general* model that uses machine learning algorithms to *learn* glioma growth patterns from real patient data, and to predict tumour growth at the voxel level, based on a classification model learned from the patient data and from voxel-specific features.

Chapter 3

Glioma Modeling Framework

As described in Chapter 2, many researchers base their work on simulating virtual tumours often because of the lack of clinical data. Access to patient data is usually under restrictions by health institutions due to confidentiality matters; because of this, the opportunity to collect such data is not always available to researchers. But for the purpose of this project, we were able to collect a set of MRI scans for a variety of glioma grades, which is crucial to the learning and prediction components of our model and to the evaluation of the results (see Chapter 4).

In this chapter, we give an overview of the framework of the proposed model including the tools and methods utilized in the several steps of the data processing pipeline and model development. Because data is an important part of this work, we also describe in this chapter how we preprocessed the MRI scans, and extracted the learning features from the processed image slices.

We collected the MRI scans at the Cross Cancer Institute (CCI) in July 2003. The database from which we obtained the brain images was relatively recent. The studies that were available to us cover only from the year 2001 to July 2003.

The total dataset we collected contains multiple consecutive scans for about 200 patients. In the dataset, the largest time interval between two consecutive scans of a single patient is less than two years. Each scan consists of several series in three different views of the brain: axial, sagittal, and coronal. The axial view is usually represented in three different modalities: T1, T1 with contrast, and T2. Each series may be either an axial, sagittal or coronal view of the patient's brain. Each series has $20-21$ consecutive slices with a $6.5mm$ slice-interval. Axial images are usually taken in the three main modalities: T1-weighted, T1 with contrast, and T2-weighted. See Section 1.4 for more detailed descriptions of MR imaging modalities and the interpretation of brain images.

Combining the slices of a series or the axial, sagittal and coronal views of the same brain provides medical staff with a 3D image of the patient's brain and helps them identify the diseased regions. Scanning the patient after a surgical procedure or during a radiation course allows medical doctors to follow-up the patient's progress and to track further tumour growth. Most patients, in particular those diagnosed with high-grade gliomas, are scanned regularly depending on diagnosis and treatment.

We carefully studied the data collection and selected a subset that is most useful to the project. The empirical data actually used in the experiments represents only a small subset of the overall 200-patient collection. We have selected this experimental subset based on two main factors: the quality of the scan, and the visible tumour growth observed between the consecutive scans of the patient. Our data subset covers a variety of patients from different age categories, and with different tumour grades ranging from low-grade gliomas to GBM. The experiments reported in this thesis were performed only on the axial views of the scans using all three modalities.

We first describe the steps involved in the data processing pipeline. Processing the data has been a large part of the framework in this thesis and is a pre-requisite for the feature extraction, and the learning and classification tasks. The processing of the data has been performed with the help of our research team, using the methods suggested in [95]. Data processing includes noise reduction and other important elements that are required for the implementation of the proposed diffusion model and that help improve system performance. The implementation of the proposed model can be summarized in three steps: feature extraction, learning and classification, and the modeling of the tumour growth volume. These three steps represent the contribution of the thesis.

The steps involved in the framework of the proposed model are outlined as follows and also illustrated in Figure 3.10 (at the end of this chapter):

1. Noise reduction is meant to reduce the effect of noise, inter-slice intensity variations, and intensity inhomongeneity in the brain imaging data sequences.

2. Registration is the process of spatially aligning the image to a template in a standard coordinate system and of warping the image regions to the template.

3. Intensity standardization is transforming the image intensities to provide a more meaningful approximation of the image regions aligned to those in the template.

4. Tumour segmentation is identifying the tumour volume and boundaries in the brain scans either manually or automatically.

5. Temporal interpolation is a method for modeling the intermediate tumour volume between two time points given the initial and final tumour volumes. This method can be utilized in standardizing the tumour volume increase throughout the dataset.

6. Tissue segmentation is separating regions of brain tissue into distinct classes that will be used in the classification and modeling tasks.

7. Feature extraction is finding voxel-level features that provide information about the regions adjacent to the tumour.

8. Classification is the step in which we use a classifier, learned from the patient data and the attributes of the voxels, to assign each unlabeled voxel in the brain volume of the test patient to the 'tumour' or 'non-tumour' class.

9. Diffusion modeling is the actual prediction of the tumour growth – given the initial tumour volume of the test patient at one time – based on the probabilistic labels predicted in the classification task. The output of this final step is a contiguous 3D volume that corresponds to the volume of the tumour at a later time.

We briefly mention from a technical perspective the image formats involved in the preprocessing of the data. The medical images we acquired from the CCI were stored in a format specific to medical tools and applications: Digital Imaging and Communications in Medicine (DICOM), created by National Electrical Manufacturers Association [34]. This is the standard image format in the medical field and is used in storing MRI patient scans. Each DICOM image file has both a header and a data file containing the stream of pixels. The header contains information about the patient, type of scan and image dimensions. We removed patient data from the DICOM header due to patient confidentiality issues. In addition, we converted the image volumes from the DICOM format to Portable Network Graphics (PNG) due to the lack of portability of the DICOM format. For this purpose, we used DicomWorks [35] to obtain PNG image slices for each of the scans by recursively processing the subfolders of the patients. The PNG files were then converted to Portable Pixel Map (PPM) image format to make the data available for feature extraction.

3.1 Noise Reduction

This is the first step in the image processing pipeline and is meant to reduce the effect of noise in the imaging data sequences. It is preferable to perform noise reduction before further processing of the data but this step is not as fundamental as other processing steps (see Sections 3.2 and 3.3), which are required for the prediction task. We use the term 'noise' to refer to inter-slice intensity variations, intensity inhomogeneity across the consecutive slices of an image volume, and misalignments of patient scans along the z axis. The input to the noise reduction step can be either the raw images or the transformed images after registration – depending on the type of noise reduction methods used in this first step. The output will be the same images but where the noise in the data, intensity variations intensity inhomogeneity, and misalignments having been reduced. In the preprocessing of our data, the inter-slice intensity variation reduction was performed before registering the images while inhomogeneity reduction and slice alignment were performed as post-processing steps to image registration since registration helps partially reduce noise in the image volumes.

3.1.1 Inter-slice Intensity Variation Reduction

Inter-slice intensity variations are observed as sudden changes in the intensity values across the consecutive slices of a scan, and depend on the acquisition protocol used in producing the imaging data sequences. As a result, it is common to observe in a single scan even-numbered slices being noticeably darker than odd-numbered slices or vice versa [97]; see Figure 3.1. Because this intensity variation effect is present in some of our data, we applied inhomogeneity reduction to the scans. Several methods can be applied in smoothing the intensity variations of MRI scans. Some of these methods are surveyed in [95]. For the purpose of our task, we used a local implementation of the weighted least squares estimation method for reducing inter-slice intensity variations [95]. This method computes the linear mapping to the median slice in the sequence from each of the two adjacent slices, transforms these slices, and then estimates the intensity mappings of their adjacent slices, until all slices have been transformed. This method has the advantage of accounting for the presence of tumours in the scans (Figure 3.1).

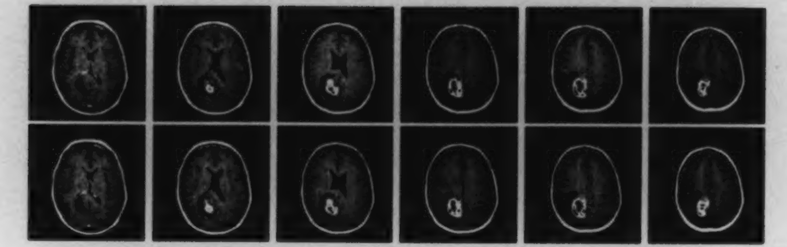

Figure 3.1: Example of inter-slice intensity variation reduction
Top: slices from same scan before reduction of inter-slice intensity variation. Note the alternating between bright and dark slices. Bottom: the same slices after reduction of inter-slice intensity variation. The images are moving from bottom to top of the patient's head, left to right.

3.1.2 Intensity Inhomogeneity Reduction

Intensity inhomogeneity, or bias field, is the corruption with a slowly varying multiplicative spatial field across the image volume, and represents one of the major problems of automated MR image segmentation. This problem is inherent to MR imaging and is caused by the limitations of the current MRI equipment patient-induced electrodynamic interactions [98]. Bias field is not always visible to the human observer, but it causes significant tissue misclassification problems when intensity-based segmentation is used.

Therefore, it is required to correct intensity inhomogeneity in the image volume prior to segmenting it into the distinct tissue classes (see Section 3.6). Several bias field correction methods are surveyed in [45, 93]. Due to the lack of real data and of studies that effectively compare such methods, it becomes difficult to quantitatively measure the performance of each of these methods. This is why it seemed reasonable to choose the intensity inhomogeneity correction feature provided in SPM [6], which is the same tool we also use for the tissue segmentation step described in Section 3.6. Intensity inhomogeneity correction was applied prior to image segmentation by setting parameters that control the correction of bias artifact.

3.1.3 Slice Alignment

Slice and angle misalignment is a problem observed in some image volumes, even in consecutive studies of the same patient (see Figure 3.2). This problem occurs when the patient's brain is being imaged as the MRI technician may start the scanning process at a different angle of the brain or at a different position along the z axis of the patient's head, as compared to previous scans of the same patient. Consequently, the same slice position in different scans of the same patient may show different brain structures. For example, the tumour or the ventricles may appear in different positions along the z axis in the consecutive scans of the same patient.

It is important to address this problem because training, prediction, and model evaluation are performed at the voxel level. Therefore, the voxels must align properly along the z axis in the successive image volumes for each of the patients. Slice and angle misalign-

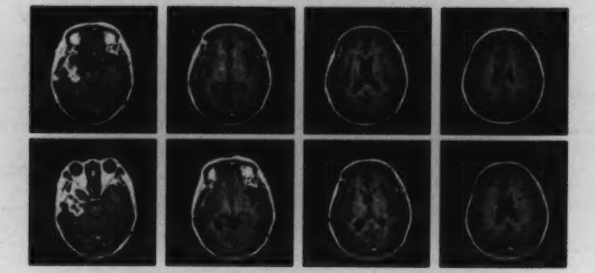

Figure 3.2: Example of slice misalignment along the z axis
Top, left to right: slices of an image volume at one time point showing brain anatomy from lower to upper brain. Bottom row: slices of an image volume of the same brain, at a later time. Images of the top row correspond to images of the bottom row but they show slightly different brain regions due to misalignment along the z axis.

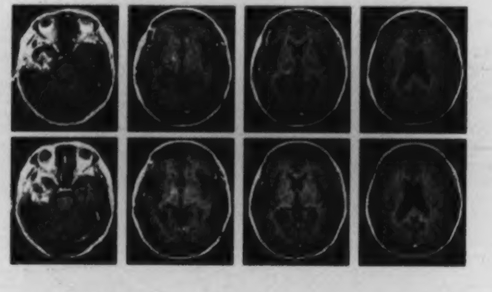

Figure 3.3: Reduction of slice misalignment after image registration with SPM
Images correspond to those in Figure 3.2. SPM's image registration to the Colin Holmes template has helped significantly reduce slice misalignment along the z axis in most cases.

ments are partially corrected after registering the brain image volumes to a template with the help of SPM's spatial registration algorithm [3, 4, 43]. In particular, misalignments may be entirely eliminated with image registration to a template if the misalignment along the z axis is only $2 - 3mm$ (see Figure 3.3). In rare cases, however, the misalignment can be as severe as a $10mm$ difference on the z axis between image volumes. This misalignment can be corrected to a large extent with image registration to a template of an individual brain (*e.g.*, the Colin Holmes template [54]) rather than an average template of several brains. In addition, semi-automated adjustment can be performed by sliding the brain images up or down the z axis by the number of misaligned slices. This adjustment needs to be performed only on segmented tumour volumes after registration and spatial interpolation. But we have not needed to apply this type of adjustment to any of the image volumes as registering the images to the Colin Holmes template [54] – a high quality average image of the brain of an individual rather than a population – helped eliminate the misalignment from the majority of the image volumes and significantly reduced severe misalignments.

3.2 Registration

Image registration is the task of spatially aligning two image volumes. A registration algorithm aligns two image volumes by computing a transformation matrix that maps one image volume to the other, the second image usually being a template. When the registration task is complete, the pixels in the transformed image have the same size and locations as the corresponding pixels in the template. Image registration consists of several steps which are performing a linear affine transformation, followed by non-linear warping and finally spatial interpolation. An accurate algorithm should not only be able to align the image volumes such that their similarity is maximal but the algorithm should also provide some degree of regularization to prevent unreasonable image deformations, in particular with the presence of tumours in the brain images. The registration steps below aim at registering a patient image to a template image where both image volumes have the same modality (*i.e.*, aligning a T1-weighted image with a T1-weighted template). But there also exists registration methods that align images from different modalities (*e.g.*, aligning T1-weighted images with T2-weighted images).

The registration algorithms are applied to the data after the inter-slice intensity variation reduction step. The output of the registration process is a transformed image that has the same coordinate system and highly similar spatial regions as the template. The registration process is fully automated and performed on three-dimensional image volumes, and does not depend on any manual recognition or segmentation of the spatial regions and landmarks in the images or the template. This automatic registration method accounts for the variety of the human brain anatomy that is obvious in the data and in particular with the presence of large tumours in the brain scans.

There are several image registration packages that could be used to accomplish the linear and non-linear registration and spatial interpolation, as described below in the rest of this section. But we ultimately chose SPM (Statistical Parametric Mapping [104]) which is a package that combines several algorithms specifically designed for the analysis of brain imaging data sequences. SPM is commonly used for processing brain images in the research of brain diseases including cancer. Statistical Parameteric Mapping refers to the construction and assessment of spatially extended statistical processes used in testing hypotheses about functional imaging data. The imaging data sequences can be a series of images from different cohorts, or time-series from the same subject. SPM processes brain imaging data sequences (or an image volume) in a special format, the ANALYZE format, such that the image volume consists of a header with dimensional information and a file containing the stream of voxels. SPM requires setting several parameters in order to obtain the best possible image registration and noise reduction. SPM provides the advantage that it is easy to vary its different parameter settings, and so it is widely employed in the processing of brain imaging data sequences. We explain in the remainder of this section the significance of each of the registration steps and the algorithm utilized in performing each step. In Figure 3.4, we show an example of image volume registration using SPM.

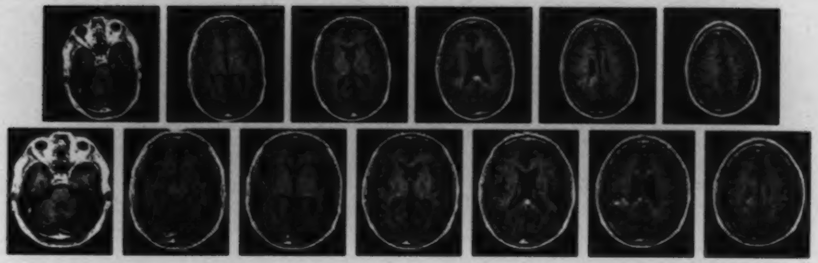

Figure 3.4: Example of an image volume before and after registration
Top, left to right: slices from the same image volume moving from the bottom to the top of the head before performing the registration steps with SPM. Bottom: slices corresponding to the top row after registration to a template, non-linear warping and spatial interpolation. Note that the input volume to the registration process is 20 slices while the output is a 91-slice volume.

3.2.1 Linear Registration

The registration of the MR image volumes to a template provides a way of linearly aligning the images within a standard coordinate system and allows the standardization of the voxels size. This linear registration step is required by our feature extraction task because the learning and classification components are performed on the voxel level. That is, when we register the images to a standard template, the classifiers can use same size voxels while the voxels in the original image may be of different sizes.

The most commonly known brain coordinate system used in brain image registration is the Talairach coordinate system [117]. The Talairach brain is the brain dissected and photographed for the Talairach and Tourneaux atlas labeled with an approximation of the functional areas in the brain. Another well-known coordinate system is the Montreal Neurological Institute (MNI) system [37, 38], which is defined based on an average template of a number of individuals, and is therefore more representative of the average brain and provides spatial prior probabilities. The MNI template was built by first manually defining landmarks on 241 normal MRI scans and by approximately matching the brains to the Talairach atlas. Then 'MNI305', the first MNI template, was made by matching 305 normal scans, with the help of a linear algorithm, to the average of the 241 brains that were already matched to the Talairach atlas. The current standard MNI template is the ICBM152, which is the average of 152 normal MRI scans that have been matched to MNI305. Another important template produced by MNI is the Colin Holmes [54], which is a template of a normal brain of an individual rather than a population. Holmes was scanned 27 times, and the scans were coregistered, averaged and matched to MNI305 to produce a high detail MRI dataset of one brain. MNI templates are included and used by the linear registration and non-linear warping algorithms in SPM [3, 4, 43]. See Figure 3.5 for an example slice of each of Colin Holmes, and the ICBM average templates.

To linearly register our dataset of MRI scans, we used the Colin Holmes template because of its high resolution and image detail. We performed image registration to the Colin Holmes template by utilizing the linear registration method provided through the SPM package [3, 4, 43]. This method has several advantages as it accounts for noise in the

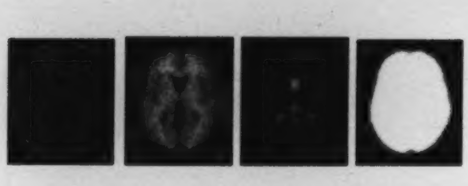

Figure 3.5: Spatial priors used in registration and feature extraction
Left to right: an example slice of each of the Colin Holmes template, ICBM T1 average template, T2 average template, and the brain mask prior.

data including intensity inhomogeneity, anisotropic voxels, image tilts, and inter-slice gaps.

3.2.2 Non-linear Warping

Non-linear warping is a step that follows linear registration, and consists of applying some deformations to the image volume so that the image would be more properly lined up with the template to which the image is registered. The purpose of using non-linear registration is to partially correct observable differences in the images by mapping spatial regions on the image to the template. Differences in the images include the head shape of the patient, the anatomical variability across patient brains, and other image abnormalities. Non-linear registration significantly increases the degree of resemblance between the image volume and the template, and allows more reasonable comparisons between the anatomical regions of the image and those of the template, in particular with the use of a template of a single individual rather than an average of a population. Another alternative is to warp the template to the patient's image but this option will not provide the advantage of correspondence between the regions in the images and the regions in a standard template. Since the learning and prediction tasks in our proposed model act on the voxel level and use data instances from several patients, it is preferable to warp the brain images to the template rather than alternatively warping the template to patients' images thus maintaining standard regions across the patients' images.

The non-linear registration algorithm we applied is included in the SPM package [104] and presented in [4], a widely-used algorithm known for its computational efficiency and performance. We have chosen this particular algorithm among several others as it also provides a high degree of regularization to correct for excessive local deformations when the image is being warped to the template.

3.2.3 Spatial Interpolation

Spatial interpolation is applied after non-linear registration and is used to compute the locations and intensity values of the pixels in the transformed image volume. Among the several spatial interpolation methods, we use high-order β-splines spatial interpolation provided through SPM [104] and described in [7]. High-order β-splines interpolation is known for its accuracy and computational efficiency as compared to other spatial interpolation algorithms; see [95]. The input to the spatial interpolation is a 20-slice image volume that has already been registered and warped to a template. The output is a 91-slice volume that has voxel size 8mm^3. The number of slices in the output volume and the voxel size are spatial interpolation parameters that can be set by the user. In our data, we generate 8mm^3 voxels (rather than $1mm^3$) to reduce the number of training instances per brain volume (*i.e.*, to increase the efficiency of the learning algorithms) while still using small enough voxels to help predict tumour growth as accurately as possible – as there is always a trade-off between efficiency and accuracy.

3.2.4 Summary of the Registration Task

In summary, image registration includes three steps applied to the MRI brain scans, to produce the information that we use later in the learning and prediction components of the proposed glioma diffusion model. These three steps are linear registration, non-linear warping, and spatial interpolation. The linear registration of a brain scan to a template is used to produce a standard coordinate system for the image volumes in the dataset. The non-linear warping performs some deformations on the image volume so that the regions in the image match more properly the regions in the template. The spatial interpolation computes the intensities and voxel sizes in the transformed image, and fills the inter-slice gaps.

3.3 Intensity Standardization

Intensity differences across the image volumes of the same and of various patients can be observed as relatively brighter or darker image volumes (see Figure 3.6). These intensity differences across patients' scans represent a problem when performing intensity-based classification. Intensity standardization is a fundamental step that allows the intensity values of the brain images to be utilized in the classification framework without accounting for intensity differences that would have required patient-specific training. The goal of the intensity standardization step is to convert the intensities of the input image volumes to an intensity distribution where the values of the intensities will provide a more meaningful approximation of the anatomical information in patient scans. This is done by converting the intensities of the input image to an intensity distribution where the intensity values will have a more meaningful anatomical approximation. But unlike inter-slice intensity variation reduction, intensity standardization is a more complicated task due to the presence of tumours and edema that differ in size and locations in the brain images of the dataset. The main methods suggested in the literature for this task are histogram-based, model-based, and template-based [137] approaches. But histogram-based and model-based methods are not suitable with the existence of large tumours and edema sections in the brain images, which will corrupt the histogram distributions and model estimations. These approaches and other intensity standardization methods are surveyed in [95] that also suggests a symmetry-based, weighted linear regression approach that can be applied to each slice of the image volume rather than computing a global factor to reduce computational cost. Since this weighted linear regression method [95] has been tested on a subset of our data collection, and is robust to the presence of tumours in the images, we selected it among other methods to perform the intensity standardization task. Figure 3.6 shows an example of the output of this method.

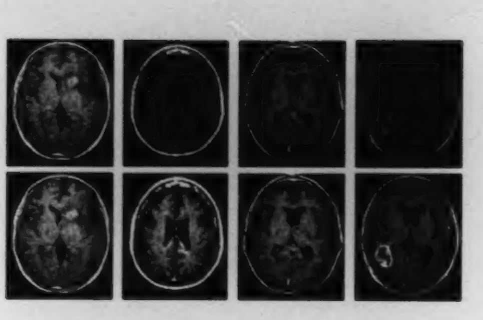

Figure 3.6: Example images before and after intensity standardization
Top: slices obtained from different image volumes for four different patients after registration to a template before performing intensity standardization. Note the intensity differences across the four slices, in particular the noticeably darker intensities of the second and fourth slices. Bottom: slices corresponding to top row after intensity standardization.

3.4 Tumour Segmentation

In order to model tumour growth, we needed first to delineate the tumour volumes from the brain images. Throughout this thesis, we use the term 'tumour segmentation' to refer to the identification of the tumour boundary or the delineation of the tumour volume on the brain scans either manually or automatically. Tumour segmentation has been the most challenging image processing problems we have encountered throughout the entire data processing pipeline. We have looked into several automatic tumour segmentation methods and compared their usability given the data we had collected. We found that some of these methods required preprocessing steps such as brain extraction (or skull stripping), which in turn involved using brain extraction tools. MEDx [80] is a software package that provides several brain image processing tools including an implementation of the skull stripping algorithm proposed in [100]. But due to the presence of the tumour in the image, brain extraction algorithms do not usually perform well and eventually remove from the image the tumour along with the skull, in particular when the tumour is significantly large. This is because the tumour and the bone may have a similar signal on the image in particular with the presence of a contrast agent and when the tumour location is near the skull. Some of the tumour segmentation methods that we have initially investigated for the purpose of our task are MR ADORE (originally developed to recognize pine trees in forestry images) [13], fuzzy clustering [20], and snakes (a method of generating contours) [52, 53]. But these methods require some preprocessing steps such as brain extraction and algorithm initialization, are sensitive to noise in the MRI scans, and therefore incapable of finding contiguous tumour regions in the image without first applying these preprocessing steps. There exists a multitude of other automated tumour segmentation algorithms including supervised and unsupervised methods, several of which are surveyed in [95]. Some of these methods are simple and rely on the contrast agent in the brain image to segment the enhancing tumour but often miss necrotic regions. Other methods are more robust and use Expectation Maximization or classification-based approaches that utilize intensity and textural features from the images (see descriptions of these methods in [95]). Overall, automated tumour segmentation

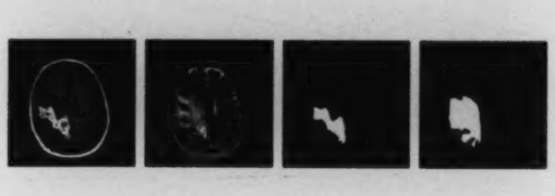

Figure 3.7: Example of segmented tumour and edema volumes

Left to right: T1-weighted image with contrast and T2-weighted image (before any processing), followed by the manually segmented tumour and edema volumes obtained from these T1-contrast and T2 slices, respectively.

is still an area that could use much more research and improvement.

Because tumour segmentation is a challenging problem that can be invested in a separate research project and because the accuracy of the segmented gross tumour volumes (GTV) is a requirement for our system, we chose to manually delineate the tumour volumes from the brain images with the help of a tool locally implemented. The annotations of the tumours for the selected data subjects were manually completed and were verified by medical experts.

The manual tumour segmentation task can be summarized in three distinct steps. First, the delineation of the tumours was done by manually drawing the contour of the tumour on T1-contrast images where the tumour is often visibly enhancing. Annotations of edema have been drawn on T2 images where fluid (water or the tumour's leaky blood vessels) enhances. Second, we extracted the segmented tumour volumes and edema volumes from the annotated images, and finally, we applied the above registration steps to the segmented volumes to obtain tumour and edema images in the same coordinate system and pixel size as for the other registered images. These registration steps are required for feature extraction, and for the voxel-level learning and classification tasks.

Manual segmentation of tumours was our best option in comparison with the existing automatic methods, but it should be mentioned that it is after all a subjective task susceptible to radiologists' agreement. That is, different radiologists may have different interpretations with respect to the determination of the tumour volume and boundaries on the MRI scan [23, 58, 78, 102]. More specifically, Mazzara et al. [78] quantified an average of 28% variation between individuals performing the same brain tumor segmentation (the variation ranged from 11% to 69%), and quantified a 20% variation within individuals repeating the task three times at 1-month intervals. Manual segmentation also includes an error margin at the tumour boundary, *i.e.*, if the same radiologist segments the same tumour twice in a row, the two segmentations will likely be different especially at the tumour boundaries. Despite these issues, the annotation of a human expert will still be more accurate than automatic methods as it ensures that normal brain regions that have a similar signal as the tumour (*e.g.*, enhancing blood vessels on T1-contrast and CSF on T2-weighted images) are not incorrectly delineated as part of the tumour volume. But manual segmentations are usually not as consistent as automated ones when a number of tumours are being segmented due to subjective judgement and the error margin at the tumour boundaries.

3.5 Temporal Interpolation

While spatial interpolation consists of computing the pixel sizes and filling the inter-slice gaps in an image volume, temporal interpolation is associated with the prediction task (see Section 4.3.3) as it provides a method of standardizing the growth increase and quantifying tumour growth among patient studies in the experimental dataset. Temporal interpolation is a method that we have investigated to provide a standard tumour volume increase across the training data instances. For example, in a GBM patient, the tumour may become twice as large over 6 months while in another patient, diagnosed with a low-grade astrocytoma, the tumour volume may only increase slightly over one year. This lack of standard volume increase among patient scans represents a problem when performing volume-based classification. Temporal interpolation provides a solution to this problem as it allows the modeling of the tumour based on a standard increase in tumour volumes across patients' scans (see Figure 3.8).

But we have ultimately eliminated the temporal interpolation step from the data processing pipeline. In this section, we define temporal interpolation, we describe the challenges involved in implementing the algorithm, and we explain the reasons in eliminating this step from the data processing and the alternative techniques that can be used in generating training data.

The input to a temporal interpolation algorithm is the tumour volumes from two consecutive patient scans. The output is a modeling of the tumour's shape at an intermediate time point between these two consecutive scans. The volume of the interpolated tumour is specified by the user (as a value that depends on a standard growth factor across patient scans), and must be larger than the tumour size observed on the first of two consecutive scans and smaller than the volume detected on the second scan. Therefore, the algorithm can be applied only if there is sufficient tumour growth between the consecutive scans of the patient. The output of the temporal interpolation will be used as the 'truth volume' from which we obtain the class labels in the prediction task and which we use in evaluating the results of the diffusion models. Since the interpolation's output will be used as *ground truth*, it is necessary that the interpolation algorithm be able to accurately model the tumour's shape and the direction of growth given the volume at an intermediate time. We provide below a brief overview of the temporal interpolation method and how we initially used it in the learning and prediction components of the proposed diffusion model.

Given the tumour volumes at two consecutive time points t_0 and t_2, our temporal interpolation method is designed to model the tumour's shape, volume and direction of growth at a particular but unknown time point within the time interval between t_0 and t_2. For example, we may have two consecutive scans at time points t_0 and t_2 respectively. But we may not have any scans for the patient at time t_1 such that $t_0 < t_1 < t_2$. Given a user-specified tumour volume v_1 that is larger than the original volume v_0 at time t_0 but smaller than the volume v_2 at time t_2, the temporal interpolation algorithm should ideally be able to model the tumour at time t_1 based on its volume v_1. The challenge in this task is to be able to implement an accurate algorithm that correctly models the unknown shape and the direction of growth of the tumour at some unknown time t_1 when the tumour reaches a user-specified volume v_1 where $v_0 < v_1 < v_2$. Note that we assume consistent tumour growth between t_0 and t_2 (based on a growth factor specified by the user), and we exclude subjects where the tumours have decreased in volume due to surgical operations or radiation therapy.

Given the properties that the temporal interpolation algorithm should have, the literature did not provide standard methods that can be directly applied to brain images and particularly to tumour volumes. We implemented a simple interpolation algorithm that allows the user to model the tumour volume v_1 given an initial and a final volume, v_0 and v_2, respectively. This implementation consists of three steps. First, we compute the Manhattan distances between the border voxels of the tumour at time t_0 and the border voxels at time t_2. We start from boundary pixels of the tumour volume at time t_0, and grow the tumour outwards recording the shortest Manhattan distance until we hit the boundary at time t_2. The second step is to compute the normalized Manhattan distances. To obtain these normalized distances, we divide the Manhattan distances by the maximum distance along the path. Finally, we compute the tumour volume at time t_1. We do this by filling in adjacent voxels starting with those that have the highest normalized Manhattan distance, until we obtain the user-specified volume v_1.

To test the performance of the temporal interpolation algorithm, we must have access to at least several instances of three scans of the same patient, such that the algorithm will model the tumour volume at the intermediate time point among these three scans. Because our data collection contains only a few instances of three consecutive patient scans that show significant tumour growth across the three scans, it was not feasible to thoroughly test the above implementation. In several cases, consecutive scans were taken at $3 - 6$ months interval; many of these did not have progressive tumour growth within these time intervals, which made it difficult to properly quantify the average performance of our implementation. Also, treatment (usually surgery and radiation) is a factor that affects the tumour growth rate and direction. In some cases, it may not be possible to accurately model the location and shape of the tumour at volume v_1 simply based on the volumes and shapes of the tumour at times t_0 and t_2.

We tested our temporal interpolation method on two patients with invasive glioblastomas; each patient has three consecutive scans where the tumour volume consistently increased. We modeled the tumour volume v_1 for each patient given the tumour volumes at times t_0 and t_2. We evaluated the results of the interpolation algorithm against the volume v_1 detected at the intermediate scan at time t_1. This evaluation of the algorithm yielded only $65 - 70\%$ accuracy.

In addition to performance concerns, applying temporal interpolation would restrict the empirical dataset to only those patients that meet a specific threshold of tumour volume increase. This is because the tumour volume increase should be standardized among patients when performing volume-based classification. For example, if the classification is based on a growth factor of 2, all tumour instances that did not show growth at least twice as large as their initial volume at t_0 will be excluded, which will result in restricting the number of data instances we could use in the training and classification tasks to only tumour instances that increased in volume by more than ($2\ v_0$), and will not represent various tumour grades in particular low-grade gliomas.

Given the restrictions imposed by the temporal interpolation method, we eliminated this step from the data processing framework. We alternatively introduced features that describe the volume increase computed from the tumour growth observed between the consecutive scans of the patient. We also obtain the class label required for the prediction task from the

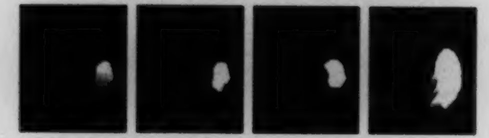

Figure 3.8: Comparison of the real tumour with temporal interpolation results

The slices shown in the figure are obtained from scans of the same patient at three different times. Left to right: the first two slices show the segmented volume of the real tumour at the first and the second time points, respectively. The third slice shows the modeled volume of the tumour at the second time point (*i.e.*, the interpolation output). The fourth slice shows the real tumour at the third time point. The interpolation output is an intermediate volume between the first and third time points, and has the same number of voxels as the volume at the second time point. We compare the interpolated volume with the real tumour at this second time point (*i.e.*, the second and third slices on this figure). Obviously, the two volumes differ in size and in shape at the slice level.

second of the two consecutive scans (see Section 4.3.3).

3.6 Tissue Segmentation

By performing tissue segmentation, we differentiate between the distinct types of tissue regions in the brain. The non-homogeneous anatomy of the brain includes bone, neural and glial tissue, and an outer membrane. Cancer tumours do not grow in all types of tissue alike. In this project, we are mainly concerned with glial tissue where tumours originate and grow. We differentiate between three types of tissue involved in tumour diffusion, which are grey matter, white matter, and cerebrospinal fluid (CSF). Tumours do not grow in CSF (*e.g.*, inside the ventricles) but they push and deform these anatomic structures.

Recent research suggests that gliomas proliferate faster in white matter than in grey matter [110] and that tumour diffusion follows white matter tracts [90]. We take these suggestions into account by including features that describe the heterogeneous brain tissue. These features may provide insight into the direction and growth rates of the tumour.

To differentiate between the brain tissue types, we segment the MRI brain volumes into three distinct classes (grey matter, white matter, and CSF). We use Ashburner and Friston's tissue segmentation algorithm [3, 5, 6] provided in the SPM package [104]. This tissue segmentation algorithm performs cluster analysis on image volumes, already registered to a template, with a modified Mixture Model and a-priori information about the likelihood of each voxel being one of a number of different tissue types. The algorithm then performs a cleanup of the partitions and finally writes three different image volumes, one for each tissue type. We preferred to use Ashburner and Friston's algorithm for its known performance in comparison with other segmentation methods such as FAST [136], a tool in the FSL package [100] that segments large brain regions, as it yielded a very high number of false positives and false negatives when applied to our data.

The output of SPM's tissue segmentation is three distinct volumes: grey matter, white matter, and CSF (see Figure 3.9).

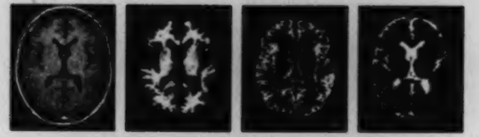

Figure 3.9: Example of segmented grey and white matter and CSF
Left to right: T1-weighted image, white matter, grey matter, and CSF segmented from the T1 image using SPM. Each of these represents one slice in a 91-slice image volume, and they all correspond to the same position on the z axis.

3.7 Feature Extraction

Feature extraction is a pre-requisite step for both the training and classification tasks. This step can only be performed after applying a series of image preprocessing steps (described in the above sections of this chapter) that consist of noise reduction, image registration to a template, segmentation of the registered images into distinct tissue classes, delineation and registeration of tumour volumes to the same coordinate system as the brain volumes. After applying these preprocessing steps, we extract image-based features from the brain images by obtaining pixel-specific intensities and labels for each image volume.

Most of the learning features that we use in the training and classification are image-based features that can be directly computed from the images. We compute tumour-specific and edema-specific features based on the labels extracted from the segmented tumour and edema volumes as described in Section 3.4. Tumour and edema labels are based on intensity thresholds obtained from the segmented tumour and edema volumes after registering and spatially interpolating these volumes to a template. We also use features local to the voxel, which include coordinate-based features and intensity-based features directly extracted from the processed images, and feature-based features computed from the difference between pixel intensities and the distances between the voxels and the tumour boundary. Neighbourhood features are image-based features and are specific to the neighbourhood of the training voxels. Patient-specific features, such as age and treatment information, are obtained from a separate database.

Section 4.5 describes in more detail the empirical feature space and the importance of the features involved in the learning model. We have categorized the feature space into five feature groups: brain attributes (anatomical locations), patient characteristics, tumour-specific features, properties local to each voxel in the learning set, and features local to the neighbourhood of the learning voxels.

It is worth noting that we have applied a brain mask prior [104], not as a feature, but as an alternative method to skull stripping. Brain masking allows us to distinguish brain tissue from the area outside of the brain boundary. Applying a brain mask ensures that we exclude from the training data voxels that are not located within the brain area (*i.e.*, voxels with zero intensity values as observed on the example slice of the brain mask prior in Figure 3.5). This brain masking method is simple and more accurate than other methods that use image intensity thresholds to exclude voxels outside of the brain region. Applying image intensity thresholds can be rather inaccurate since several brain regions (*e.g.*, CSF) may have a similar signal to regions outside of the brain border (*i.e.*, the skull). By using the brain mask, we can also limit the number of training instances to only voxels within the brain area, as only these voxels provide useful information for the learning model. In a similar fashion, we also limit testing instances to only voxels within the brain region. In addition, the application of a probabilistic brain mask has the advantage of replacing explicit skull stripping methods, which often have their own segmentation error margins. To remove non-brain voxels, we used the brain mask prior probability, provided in SPM2 [104], derived from the MNI305 average template [39], then re-sampled and smoothed as for other SPM priors (see Figure 3.5).

3.8 Classification

After extracting and computing the features from the imaging data sequences, we use the feature set in training a machine learning algorithm that would later predict whether a voxel will become 'tumour' or 'non-tumour' based on the features of this voxel.

In the training phase, the input to the learning algorithm represents a large dataset of several brain volumes; each volume consists of thousands of voxels described by a number of features and a voxel label obtained from the segmented tumour volume (see Sections 3.4 and 3.7). The output of the training phase is a model learned from the training features and that will be used in predicting the labels of the testing voxels. This learned model is the input to the testing phase. The output of the testing phase is the predicted class labels for the voxels of a new unlabeled brain volume, based on the features of these voxels.

For most classification algorithms, the testing phase is often computationally efficient but learning the model can be very costly. Therefore, it is preferable to limit the amount of training data to instances that will add important information to the learned model. Techniques used in selecting the training data include random and selective sub-sampling where we exclude a subset of the data either randomly or based on specific criteria, respectively. In random sampling, some training information is lost. Selective sub-sampling will likely not lead to loss of information to the same extent as with random sampling, but might produce biased classification patterns where the learned model is not always capable of distinguishing ambiguous data instances or data points that intersect between the classes.

Chapter 4 provides a more complete illustration of the training and classification components as part of the prediction task, and further describes the data, the sub-sampling techniques, and features involved in learning the diffusion model.

3.9 Diffusion Modeling

After we predict the voxel labels for the unlabeled data instances, we perform the final task in the glioma growth modeling framework, *i.e.*, the actual modeling of the tumour growth. To model the tumour growth, we apply a diffusion algorithm that takes as input a tumour volume (of a given patient at one time) and the voxel labels predicted by a classifier (see Section 3.8). The diffusion algorithm estimates the shape of the tumour once it is some k voxels larger. The algorithm starts by iterating sequentially around the active border of the tumour, considering 'eligible' voxels that are likely to become 'tumour', as predicted by the classifier. The diffusion model terminates when the tumour reaches a user-specified target volume, *i.e.*, after adding k voxels to the input tumour volume, or if the algorithm is not able to add more voxels to the tumour volume. For testing, we set this target volume to be the same volume observed on a later scan of the patient. The output of the diffusion algorithm is a prediction of the growth of the tumour when it reaches a specified volume. In other words, the output is a contiguous 3D tumour mass that has the same volume detected on a later scan of the test patient.

We describe more extensively the details involved in the implementation of the diffusion modeling task in Section 4.6. We also describe the implementations of the standard models against which we have evaluated our proposed approach.

3.10 Known Problems

It is worth mentioning that while real patient data is the foundation of the work in this thesis, the use of clinical data introduces issues some of which remain unresolved. In this thesis, however, all experiments and result analyses are solely based on the use of the MRI brain scans from which we obtain the learning features for the training and classification tasks.

In spite of the many steps involved in the image processing pipeline, there will still be noise in the data, which may somewhat affect the accuracy of the proposed model. But in this case, we will not have a quantitative measure for the loss of accuracy caused by noise in the data. That is, we may visually detect errors in some of the processing steps (*e.g.*, image registration, spatial interpolation and tissue segmentation), but we do not have a method to quantify the error margins and the loss of accuracy in terms of numerical measures. We have observed, however, that these errors are at the voxel level.

We outline the issues associated with each of the data processing steps. First, the noise reduction task can only be used to reduce noise in the data to a large extent but not completely eliminate it. After we register the images to a template, using SPM [104], most images are usually registered accurately to a standard coordinate system and should have similar approximations of the anatomical locations. But a few image volumes are originally taken at a different angle and therefore do not perfectly align on the z axis with the rest of the scans even after registeration to a standard template and after the spatial interpolation step. This leads to anatomical structures appearing in these scans while not in others along the z axis. The misalignment along the z axis is estimated to be at most $1 - 3$mm after processing the image volumes, and cannot be further corrected given the current image registration techniques.

Other registration errors include head tilts that may not be completely corrected and spatial interpolation artifacts. Interpolation errors often occur when processing slices that have very different properties at the same location. This problem is particularly more apparent when producing tumour and edema volumes, in which case the registration algorithm may not interpolate the tumour properly resulting in tumour regions overlapping with brain regions where tumours cannot grow (*e.g.*, the petrous bone). Other interpolation erros include artifacts that can be observed as sudden bright intensities or straight lines on the output image resulting in non-smooth tumour volumes across the slices of the image volume. This artifact is even more pronounced if using nearest-neighbour interpolation algorithm but mostly eliminated with the high-order β-splines interpolation approach. Both interpolation methods are available through SPM [104]. Though we have not measured the error margins in terms of numerical quantities, we have visually observed that these errors are at the voxel level.

Overall, SPM [104] was significantly easier to configure than other existing software packages that process brain scans. Also, SPM is widely recognized and used, and was even integrated in other tools such as MEDx [80]. But thus far, we still need to define and use a metric in order to quantify results of different software packages. Though SPM is known to be more robust in processing the brain imaging data sequences, value judgements were based on visual inspection of the output image volumes obtained using a variety of parameters.

In this thesis, experiments have been based on axial slices though the coronal and sagittal studies that correspond to the axial images were also available. Processing coronal and sagittal views yielded results that were less accurate than those obtained from registering axial data. While the current scope of the thesis includes only work performed on axial images, it may be worth using coronal and sagittal views (see Figure 1.4) to produce a 3D brain structure that would perhaps contribute to solving some of the data processing and research problems including the design of an accurate temporal interpolation method.

3.11 Summary of the Diffusion Model Framework

In this chapter, we discussed the sequence of steps of preparing the brain imaging data sequences and using the information in these images to model tumour growth. We conclude this chapter with the following diagram which provides a summary of the model framework including the data processing steps, and the contribution of the thesis represented by the feature extraction, the classification and the diffusion tasks:

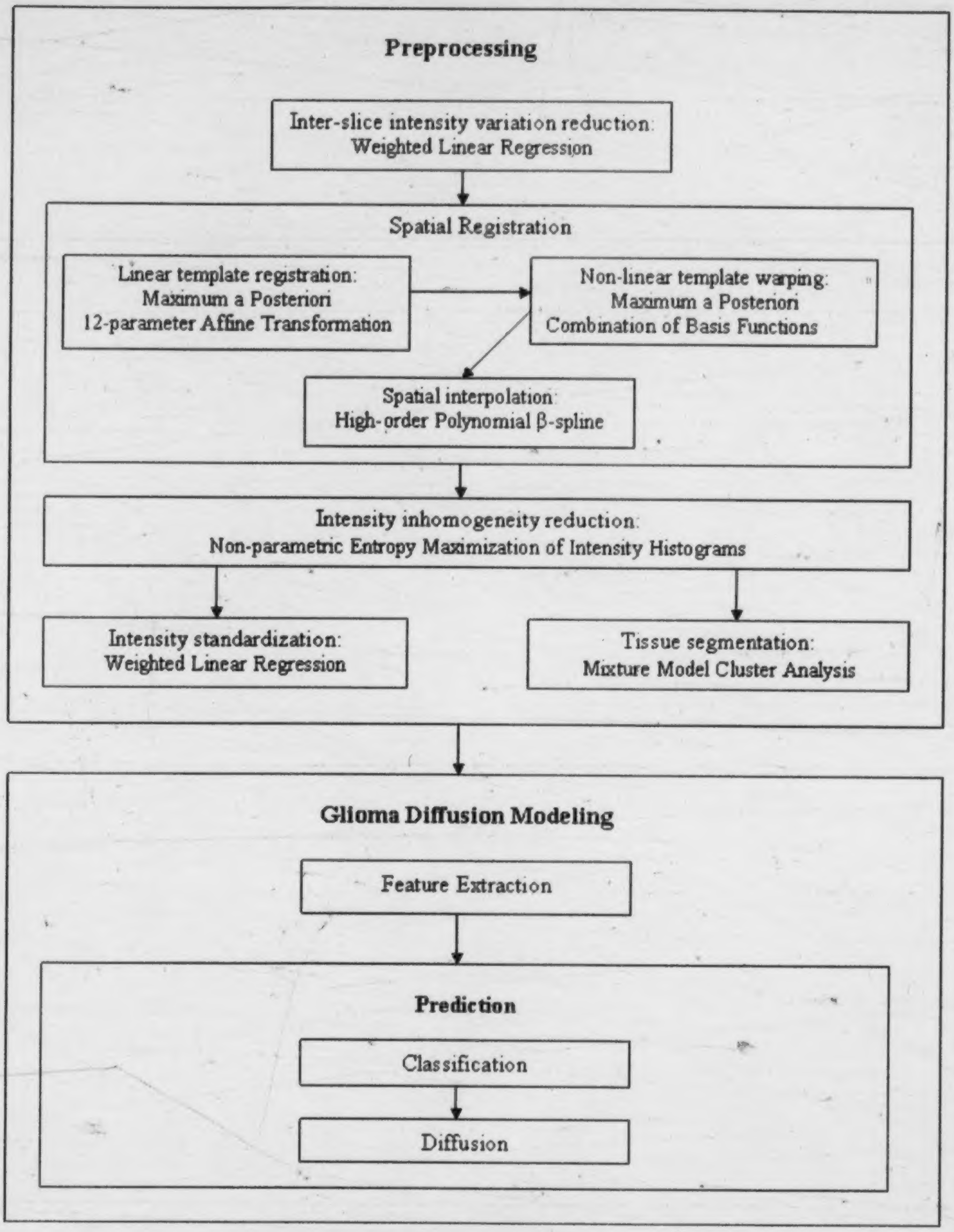
Preprocessing
Inter-slice intensity variation reduction:
Weighted Linear Regression
Spatial Registration
Linear template registration:
Maximum a Posteriori
12-parameter Affine Transformation
Non-linear template warping:
Maximum a Posteriori
Combination of Basis Functions
Spatial interpolation:
High-order Polynomial β-spline
Intensity inhomogeneity reduction:
Non-parametric Entropy Maximization of Intensity Histograms
Intensity standardization:
Weighted Linear Regression
Tissue segmentation:
Mixture Model Cluster Analysis
Glioma Diffusion Modeling
Feature Extraction
Prediction
Classification
Diffusion

Figure 3.10: Overview of the framework

The framework of the proposed model consists of two main components: the preprocessing of the MRI scans and the prediction of glioma growth. The prediction component represents the contribution of the thesis.

Chapter 4

Predicting Glioma Diffusion

This chapter provides a general background discussion of some of the existing approaches to glioma diffusion modeling, which motivates the supervised learning task, and a brief introduction to our method. Section 4.2 defines the prediction task, and describes the decisions made in the implementation process. Section 4.3 describes the training data and illustrates how we resolved data-related issues that have an implication into the definition of the prediction task. Section 4.4 describes the training and classification components involved in the prediction task. Section 4.5 outlines the empirical feature space and describes the significance of the type of features used in the learning component. Section 4.6 describes the implementations of the three models: CDM, UG and GW. The chapter concludes with a summary of our work.

4.1 Introduction to Glioma Modeling

This section presents a brief background discussion of some of the existing approaches to glioma modeling. This discussion is intended to motivate the use of supervised learning, which represents the foundation of our proposed model, followed by a general description of our implementation.

4.1.1 Background Discussion

Most models presented in the state-of-the-art literature simulated tumour growth using mathematical and statistical approaches (see Chapter 2). Some of these models merely simulated virtual tumour diffusion without initializing the simulations from clinical data. Other models are simple and made assumptions with respect to the factors involved in tumour growth regardless of the truthfulness of these factors in comparison with clinical diffusion (as described in Section 2.2.1). Statistical studies performed their analyses on clinical data obtained for a number of patients with treated and untreated gliomas (Section 2.2.3). Other models made use of clinical data to initialize and evaluate the simulations, *e.g.*, Clatz et al.'s model [21] which is a patient-specific model, and some of the research work by Swanson et al. [113] that uses clinical data to predict the growth of GBM tumours, the most common in brain cancer.

Fairly recent is the idea of involving supervised learning into tumour modeling and few are the models that can be generalized (or extended) to predict the growth of low-grade and mixed gliomas (*e.g.*, low-grade astrocytomas and MOA); in particular all models are specifically designed to originally predict the growth of GBM only. Overall, learning models are more recent; *e.g.*, Zizzari's treatment planning module [138] includes a learning component using Neural Networks to predict the irradiation volume based on mathematical approximations of the treatment volume at the macroscopic scale.

In general, researchers have more commonly simulated glioma diffusion using mathematical approaches, as opposed to using learning and classification, since mathematical models tend to be simpler and often involve a small number of parameters that could be used at once. But these approaches tend to make assumptions about tumour growth factors, and have not contributed an accurate, general model that can *learn* glioma diffusion patterns from existing patient data. Obviously, understanding glioma diffusion and discovering factors involved in clinical diffusion are not trivial tasks.

4.1.2 Motivating the Supervised Learning Task

In recent years, machine learning research and applications have received significant attention because of their latest contributions as well as their potential contributions to the different areas in Science and Medicine where learning and prediction have already been used in a wide variety of tasks. In the medical field in particular, machine learning classifiers have been applied in research topics including but not limited to disease diagnosis [105, 106], studying behavioural patterns in diseases [74], and predicting drug resistance versus treatment response [71, 87, 103]. Classifiers have also been employed in detecting tumours on medical images [95], studying and modeling tumour diffusion, and planning treatment [138] as well as in numerous other health-related areas.

In this thesis, we examine the use of machine learning algorithms in modeling glioma diffusion. We believe that supervised learning is a promising research direction that would enable us to model glioma diffusion more accurately than with existing approaches. Current glioma growth models are relatively simple, and often consider only one or two aspects of the problem (see Chapter 2). In particular, medical experts need a far more precise model than provided in the state-of-the-art literature, to help them specify the treatment volume rather than irradiating a generic 2cm margin around the detected tumour, which often results in recurrence and serious side effects.

The use of supervised learning is motivated by the advantage of capturing key information represented in the training data and the feature space. This information is crucial to the learned model, and will have a direct role in the recognition of glioma diffusion patterns and in the accuracy of the prediction task.

4.1.3 Proposed Approach

Our approach consists of predicting tumour growth, given an initial tumour volume at one time, by estimating the volume once the tumour is k voxels larger. Performing this prediction requires learning a classifier from data provided for 17 patients and from a feature set that describes characteristics of the patient, the tumour, and the regions adjacent to the tumour. The learned model is then used in classifying each voxel into either of two classes: the 'tumour' class or the 'non-tumour' class.

The diffusion model takes as input the voxel labels, as predicted by the classifier, and iterates around the active border of the input tumour volume. The algorithm updates the tumour-healthy border by possibly adding candidate voxels to the tumour mass based on the voxel's probability (*i.e.*, the probability that the voxel will become tumour as predicted by the classifier) and on the number of tumour voxels adjacent to the candidate voxel. In every iteration, the border of the tumour becomes larger and the algorithm increments the total number of transformed voxels. The simulation finally terminates when the total tumour mass reaches a specified final volume. To test the model, we use a pair of tumour volumes for each patient, such that the volumes correspond to two consecutive scans. We initialize the model from the volume at the first time point, and we allow it to grow the tumour to the volume corresponding to the second time point. Model results are evaluated in terms of precision and recall [12], and are compared to the conventional uniform diffusion (Section 1.2.2) and to a simulation of the model in [110].

The rest of this chapter defines the prediction task including the learning and classification components. It describes how we generated the training data given the data content and limitations, and lists the features used in the learning component.

4.2 Defining the Prediction Task

In this section, we define the prediction task, and we describe how we approached the problem of glioma growth modeling from a machine learning perspective. This is followed by an overview of the prediction framework, which represents the remainder of the work described in this chapter.

4.2.1 Task Definition

Glioma growth prediction has been one of the many challenging problems in cancer research. It has often been difficult to determine what factors play a role in tumour growth, its capability of invasion, and recurrence. Because of this, many researchers have dedicated their time and efforts to understanding tumour growth patterns and to discovering the factors contributing to tumour diffusion. From a medical perspective, many questions remain unanswered to this day with respect to the genetics of malignant cells and their capability of invasion. More recent research claims that malignant diffusion follows white matter tracts [90] and that tumour growth rates may be influenced by the heterogeneity of brain tissue [110]. But the truthfulness of these claims remains questionable to some extent, and it may take a few more years, or perhaps decades, of research before medical experts know for certain what factors play the most significant role in glioma diffusion.

From a machine learning perspective, we first need to answer the question, "What are we predicting?", for which the answer may seem elusive at first because of the complexity and vagueness of the problem definition itself. To understand the glioma diffusion problem and to provide a specific definition of the prediction task, it is required to identify three main components that have a direct involvement into the task definition:

- Factors that are believed to play a role in glioma diffusion, according to contemporary medical knowledge

- Data content and limitations

- Definition of *abnormality* – *i.e.*, regions we identify as 'tumour'

Our definition of the prediction task depends on glioma diffusion factors obtained from contemporary medical literature. These factors are represented in the model in terms of the feature space that we describe in Section 4.5, and include patient information, attributes of the brain anatomy, and properties of the tumour. The task definition also relies on the type of data we collected, and is limited by the data issues that we describe in Section 4.3. Furthermore, the prediction task is based on what we define as *abnormal* regions on the MRI scans. Tumour volumes are mainly obtained by segmenting the enhancing regions, as detected using T1-weighted images with contrast.

Given the diffusion factors, data limitations, and the definition of abnormality, we initialize the prediction task from the tumour volume detected on the patient's scan at one time, and we predict *where the tumour will be* when it reaches a specified final volume. We assume that occult malignant cells are responsible for further glioma diffusion and for recurrence after surgical procedure or radiation therapy. The challenge in the problem of glioma diffusion modeling is to find out where the occult cells may have infiltrated in the adjacent tissue in order to eradicate them, even though these cells may not be detectable yet on the MRI brain scan.

In the supervised learning task, we first *learn* glioma diffusion patterns based on the training data and the features describing brain regions that are likely to become diseased (see Section 4.5). The input to the training task is a set of pairs of labeled brain volumes; each consisting of a few thousand voxels. The output is a learned model based on the training data. A classifier uses this learned model to make the prediction based on what it 'recognizes' as abnormal regions and the attributes of the unlabeled voxels. The output of the classification is a label for each voxel (*i.e.*, labeled as 'tumour' or 'non-tumour'); see Section 4.4. A diffusion algorithm uses these voxel labels to model the glioma volume of the testing patient by assigning voxels adjacent to the active tumour border as either *normal* or *diseased*, based on the probabilities predicted by the classifier (see Section 4.6).

This definition of the prediction task does *not* include treatment planning specifications, *i.e.*, our model is not designed to decide where to irradiate the patient. Instead, the prediction task is about the *modeling* of glioma diffusion in a way that would enable us to understand some of the factors that have a role in glioma growth and to help medical doctors specify more properly the treatment boundaries for radiation therapy. This new specification of the treatment volume (of arbitrary geometrical shapes) may help more precisely target cancer regions while sparing normal tissue if possible, as opposed to irradiating a generic, uniform 2cm boundary around the GTV.

4.2.2 Overview of the Prediction Framework

To predict tumour diffusion, the model is initialized with a tumour volume, of a given patient at one time, and estimates the shape of the tumour when it grows by some k voxels.

We predict these k voxels by first learning from patient data a probabilistic classifier which we use to predict a label ('tumour' or 'non-tumour') for each voxel.

As a pre-requisite to the learning and classification, we must resolve data-related issues in order to generate the training data for a total of 17 patients and we also extract the learning features from the MRI scans. The input to the learning stage is the labeled voxels obtained for 16 patients and we test on the 17^{th}, that is, we use $n-1$ patients for training and the n^{th} patient for testing with cross-validation. Then, we apply the diffusion algorithm to each of the 17 patients. The output of this first stage is the learned model, used in the classification of the unlabeled brain. The output of the classification stage consists of voxel labels for the testing patient based on the features of these voxels. Each voxel is associated with a probability as predicted by the classifier. The higher this probability, the more likely this voxel will become 'tumour'.

We apply the diffusion model to the voxels of the testing patient. The model takes as input the voxel labels and iterates around the border of the tumour volume detected on the patient's scan at the first time point. The diffusion algorithm may add 'eligible' voxels to the tumour volume depending on the probability of the voxel, as predicted by the classifier, and on the number of tumour-voxel neighbours adjacent to the candidate voxel. The algorithm terminates when the tumour reaches the volume detected on the patient's scan at a later time (*i.e.*, the volume used as ground truth).

The remainder of this chapter describes in more detail the steps of the prediction framework, and the components involved in each step. Section 4.3 discusses the data-related issues, and how we approached each of these issues as they have a direct implication into the definition of the prediction task. Section 4.4 presents the learning and classification stages, the classifiers used in the experiments, and the training and testing data folds. Section 4.5 describes the feature space that will be used in the experiments, and how these features were obtained from the MRI scans. Finally, Section 4.6 describes the three diffusion models, CDM, UG and GW.

4.3 The Training Data

In this section, we describe how we generated the training data from the MRI scans, which is a pre-requisite step to specifying the learning and classification tasks. We also describe the data-related problems and how we addressed each of these problems as they have a direct implication into the definition of the prediction task, as mentioned in Section 4.2.1.

Data-related issues that emerged while processing the MRI scans, to generate the training dataset for the supervised learning task, include noise in the images, variation in tumour volumes across studies of the same and of different patients, variations in time intervals between the scans, and the class imbalance between the 'tumour' and 'non-tumour' classes.

4.3.1 Data Noise

Noise in the image volumes, inter-slice intensity variation and intensity inhomogeneity, required additional preprocessing in order to make the data usable. This was generally dealt with in the image processing pipeline as a pre-requisite to the machine learning component (Chapter 3). But it was still required to exclude data subjects from the training dataset since these subjects were not usable due to surgical cavities visible on the brain images, or other MR artifacts that could not be corrected with current image processing algorithms. This has imposed a restriction on the number of subjects (or patients) we can use in the supervised learning task.

4.3.2 Class Imbalance

The class imbalance problem is not uncommon to machine learning applications. In particular, when using real patient data such as the medical imaging data sequences used in this thesis, rather than synthetically-generated data. In such case, data bias and class imbalance are usually inevitable.

In our data, the class imbalance problem is caused by the difference between the number of normal voxels (*i.e.*, voxels that represent the 'non-tumour' class) versus the number of tumour voxels (*i.e.*, the 'tumour' class) in the brain image. Because the tumour voxels represent only a small percentage of the voxels on the MRI scan, the 'tumour' class is always far smaller than the 'non-tumour' class. Besides, the total number of training instances (combining both 'tumour' and 'non-tumour' classes) provided to the classifier can be in terms of a few millions, only $10 - 15\%$ of which belong to the 'tumour' class. Therefore, using the entire brain volume in the learning task would result in extremely large training datasets. These large sets have two disadvantages. First, the number of positive instances (*i.e.*, the 'tumour' voxels) will represent only a minor percentage of the total number of training instances. Second, training the classifier on large datasets is rather inefficient and may not necessarily yield higher accuracy.

Since most current machine learning algorithms are not designed to handle class imbalance issues, it is required to perform at least some preprocessing, or sub-sampling, on the training data before presenting it to the classifier.

For the purpose of the prediction task, we do not intend to obtain highly-balanced classes with a $1 : 1$ ratio, though we performed initial experiments with a $1 : 1$ class ratio. In initial experiments, we included all 'tumour' voxels and non-randomly sub-sampled the 'non-tumour' class by skipping a number of voxels in every row of voxels around the tumour border in a way that would provide almost equal class sizes. This sub-sampling technique resulted in non-contiguous training volumes, and did not properly represent regions adjacent to the tumour since voxels that are distant from the tumour border were also included in the training set. Instead, we generate the training data in a way to allow a proper representation of the regions of interest (*i.e.*, regions that are adjacent to the tumour and that are more likely to become diseased). To provide a reasonable representation of the positive data instances, we train on the voxels in the set difference between a pair of tumour volumes at two different times for the patient. To obtain the voxels in this set difference, we include in the training dataset all the 'tumour' voxels as well as a 2-voxel border from the 'non-tumour' class to account for segmentation error at the tumour boundary.

We obtain the class label from the second of two consecutive patient scans. The 'tumour' class represents the regions that will become tumour at the second time point excluding the visible tumour detected on the scan taken at an earlier time. The non-tumour voxels are adjacent to the border of the truth volume (*i.e.*, the tumour at the second time point). Graphically, the combination of the tumour and non-tumour voxel classes form a

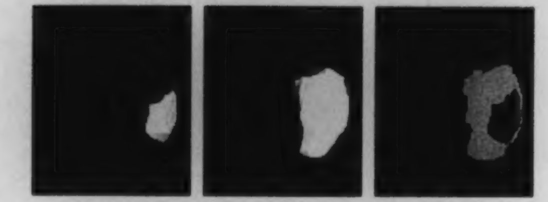

Figure 4.1: Example voxel regions at the slice level, used in training the classifiers
Left to right: the segmented tumour volume at one time, the volume at a later time (the truth) for the same patient, and an example image slice of the training data which excludes the tumour at the first time point and includes a 2-voxel border around the truth. All 3 slices correspond to the same position along the z axis.

contiguous 3D volume around the initial tumour (*i.e.*, the tumour at the first time point), see Figures 1.1 and 4.1.

This sub-sampling of the data does not result in $1:1$ ratio classes since the number of positive data instances will represent a larger percentage in the training data. This percentage is not a fixed value but depends on the initial and final tumour volumes detected at the two consecutive studies used in the training. This sub-sampling method ensures a proper representation of the regions of interest, which are voxels adjacent to the tumour, and are more likely to become tumourous as compared to the rest of the brain. Also, by sub-sampling the training data, we reduce the number of training instances provided to the classifier; hence increasing the efficiency of the learning process.

The testing dataset is generated from the voxels located within the brain border on the MRI scan (*i.e.*, the entire brain volume) but exclude the initial tumour at the first time point.

4.3.3 Time-interval and Volume Variations

Another problem encountered while processing the data is the tumour volume variations across patients and the time interval variations between studies of the same patient and studies of different patients. As a result, not all patient studies were usable even though we have collected a reasonably large dataset.

'Time-interval variations' refer to the inconsistency in time intervals elapsed between the scans of the same patient or of different patients. For example, the first two consecutive scans of a patient may be three months apart while the interval between the next two scans could be one year.

'Tumour volume variations' refer to lack of a standard growth rate or volume increase across patient scans. This means that the tumour volumes in different patients, or different studies for a single patient, may range from very small cell masses to large, invasive tumours. The variation in growth rates may be due to the effect of surgical and radiation treatment and to other diffusion factors whether these factors are presently known or unknown to the medical world. For example, one of the known factors is the tumour grade. Low-grade tumours do not significantly increase in volume over a few months, while high-grade tumours tend to be very invasive in a short period of time. Because our data collection contains various tumour grades at multiple time points, the tumour volume increase, among patients, ranges from slow-growing tumours that do not visibly grow over a year or more to aggressive tumours that become significantly large over a short period of time (typically 3 – 6 months). Another factor is the effect of treatment on the growth rate variation. Some patients may not respond to treatment. In other patients, the tumour may even show visible shrinkage on the next consecutive MRI scan of the patient because of treatment or due to a post-surgical decrease in fluid or edema surrounding the tumour.

Patient studies that showed visible decrease in the tumour mass and those that showed no growth at the next time point were excluded from the working dataset. A minimum of two consecutive studies per patient were required to perform the learning and prediction tasks. But due to the data limitations mentioned above, the number of patients involved in this thesis was only a small subset of 17 patients from the total amount of data collected. We were able to use, however, multiple studies for patients with more aggressive tumours in which there was visible tumour growth at each time point in spite of the relatively short time intervals between these studies. We were also able to include several low-grade tumours that showed detectable growth across the scans of the patient. The patients are from different age groups, 16 of which are adults and a 13-year old patient.

To address the problem of time-interval and tumour volume variations, there are three potential options to consider: time-based prediction, volume-based prediction using a standard growth rate, or volume-based prediction independent of the time interval and the volume increase between the patient's scans.

The first option is to predict glioma diffusion based on time intervals (*i.e.*, to answer the question "Where will have the tumour spread 6 months later?"). But given the data and the characteristics of the problem of glioma diffusion modeling, it is not feasible to implement the model based on this option, in particular that model evaluation will require a standard of reference (*i.e.*, a truth volume) to compare to the predicted volume.

The second option is to predict glioma diffusion based on volume increase (*i.e.*, "If the tumour is currently 5000 *cu* voxels, where will it have grown when it *doubles*?"). This second option was identified as a more reasonable task than time-based prediction as it provides a way of evaluating the predicted tumour volume against a standard of reference (*i.e.*, an interpolated volume or the real tumour volume from the next patient's scan). Also, volume-based modeling provides the advantage of comparing the predicted tumour volumes with those produced by other standard methods. With volume-based prediction, the learned model will find all voxels adjacent to the tumour border that will have become malignant when the tumour has reached a specified volume regardless of the time interval elapsed since the patient's previous scan. This requires providing the machine learning algorithm with a training dataset that depends on a standard volume increase or growth rate among the patients' instances.

One way to model tumour diffusion based on volume increase, in spite of different growth rates among patients, was to use a temporal interpolation method in the data processing. Temporal interpolation is meant to provide a consistent growth rate (or volume increase) throughout the dataset (as described in Section 3.5). This is done by interpolating between two time points and by modeling the tumour, based on a user-specified growth rate, at an intermediate time between these time points; as shown in Figure 3.8. However, we encountered two main challenges when using our locally developed temporal interpolation algorithm. First, this algorithm was relatively simple, did not yield sufficiently accurate results, and could therefore have a negative impact on the accuracy of the proposed diffusion model as we would obtain CDM's parameters and evaluate its prediction based on the results of the temporal interpolation algorithm. It is difficult and may be even impossible to accurately model the intermediate tumour shape and location by simply interpolating between two tumour volumes (see Section 3.5). Second, using temporal interpolation further limited the dataset to tumours that have a higher growth rate than the user-specified rate (see Section 3.5 for more detail). Therefore, the usage of temporal interpolation was restricted to patients with significant tumour volume increase, which would ultimately result in further limiting the experimental dataset to cases that the interpolation method can be applied to. Because of these challenges, the interpolation method was not used in the data processing. In fact, the development of a solid, accurate temporal interpolation algorithm that would not impose such restrictions on the dataset may require the time and efforts that could be invested into a separate research project.

Investigating the first two options leads to the implemention of the glioma diffusion model based on the third option, which is a volume-based prediction independent of the time interval and the growth rates of the tumours across patients' studies. This option is reasonable given the data limitations and the lack of a standard growth rate across patients' studies. This implementation requires that we provide to the learning algorithm information specific to the tumour volume since there is no standard growth rate across the training instances. In this case, we model tumour growth based on volume difference such that the diffusion algorithm is required to add to the original tumour (*i.e.*, at the first of two consecutive time points) a number of voxels, $\delta V = V_2 - V_1$, where V_2 is the final tumour volume detected on the second time point and V_1 is the volume at the first time point. For example, if $V_1 = 5000$ voxels, and we need to predict 'where' the malignant cells are when the tumour will have increased in volume by 25%, that is when the tumour reaches a volume $V_2 = 6250$ voxels. In this case, the diffusion model should predict the volume increase $\delta V = 1250$ voxels, *i.e.*, the set difference between V_1 and V_2.

The implementation of the model based on the third option does not require using the temporal interpolation step; hence preventing the accumulation of errors resulting from this step. Instead, we incorporate in the learning task features that are specific to the tumour volume (*e.g.*, the growth rate and the tumour volume increase); see Section 4.5.

4.4 Learning and Classification

After we specified the prediction task, and studied the data content and limitations, we identified the learning and classification components. In this section, we present a description of the 2-stage prediction task, and we explain how we generated the data folds used in each of the learning and classification components. We also describe the machine learning algorithms used in learning glioma diffusion patterns and in predicting tumour growth for an unlabeled brain volume.

4.4.1 Task Description

In Section 4.3, we have described the data-related issues that have a direct role in defining the prediction task. We now present a more detailed description of the components involved in this 2-stage task.

The first stage is the training process which involves building a classifier from a large number of data instances obtained for 17 patients. Each data instance represents a single 8mm^3 voxel on the brain image. Training voxels are retrieved from non-enhancing regions adjacent to the tumour, on T1-weighted images with contrast, at an initial time point. Many of these voxels will appear diseased on later scans for the patient (*i.e.*, these voxels are labeled as 'tumour' in the training data). The number of training voxels obtained from the MRI scans of a total of 17 patients is approximately $\frac{1}{2}$ million after applying the sub-sampling method described in Section 4.3.2. Note that this sub-sampling method significantly reduced the number of training instances as the total number of 8mm^3 voxels in a single brain volume is about $350,000$. By retrieving the training voxels from regions adjacent to the tumour on patients' images (rather than using the entire image for training), we are able to more properly represent the regions of interest, *i.e.*, voxels that are more likely to become 'tumour'. Since the training stage can be very computationally costly, this sub-sampling method also helps reduce the training time. Training time is approximately $30-40$ minutes when learning a Naïve Bayes or Logistic Regression classifier on the $\frac{1}{2}$ million data points. But when using Support Vector Machines, the training stage can be extremely costly, which requires further sub-sampling the training dataset by including only every 15^{th} or 25^{th} data point depending on the size of the feature set being used (see Chapter 5). This further sub-sampling reduces the training set to only $40,000$ data points or less. Note that sub-sampling is not intended to provide $1:1$ ratio classes but to represent more properly voxels of interest in the training data and to reduce the size of the dataset. The output of the training stage is the learned model based on the information content in the data and on the learning features.

In the second stage, the learned model is used in classifying the unlabeled voxels of a new patient, based on the features of these voxels. The output of the classification stage is an assigned label for each testing voxel, such that a voxel can be assigned to either the 'tumour' or the 'non-tumour' class, and a probability corresponding to the labeling of the voxel, as predicted by the classifier.

4.4.2 The Classifiers

Classifiers we considered in the prediction task include Naïve Bayes [36], Logistic Regression [68], Multilayer Perceptron [132], Support Vector Machines (SVM) [60, 88], and Decision Trees [91]. Most of these algorithms are provided through WEKA's implementations [132], which automatically discretizes the attributes in the feature space. No scaling was used; instead, the numerical values for the features were provided to the classifiers as computed by our feature extraction functions.

Given the characteristics of the data, we expect that probabilistic classifiers and algorithms that linearly classify the data into distinct binary classes, will provide more explainable results (*i.e.*, a voxel is more likely to become tumourous given that certain conditions hold – that is, if the voxel has features relevant to 'tumour' voxels) than other complex classifiers, *e.g.*, Neural Networks. In addition, given the size of the training dataset, we choose algorithms that have less costly training functions. We also account for the problem of class overlap, *i.e.*, in the training data, some voxels have the same characteristics (very similar feature values) but different class labels. Many of these voxels fall in edema regions where it is more likely that cancer cells infiltrated, but enhancing tumour masses may not yet show on the MRI scans.

Because of training data issues (*e.g.*, class overlap, characteristics of the training voxels, overfitting) and for efficiency reasons, we used in the final experiments (Chapter 5) three different classifiers to obtain a prediction probability. We then used a threshold to exclude from the 'tumour' class voxels that have a lower prediction probability than the threshold. CDM results obtained with different classifiers are presented and further discussed in Section 5.3. The three classifiers used are:

Naïve Bayes

The cornerstone of Naïve Bayes (NB) classification is the *Bayes theorem*, see [81], which provides a way to calculate the posterior probability $P(Y|X)$ of a hypothesis Y, given the training data X, from the prior probability $P(Y)$ that denotes the initial probability that hypothesis Y holds, together with $P(X)$ and $P(X|Y)$, see Equation 4.1.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \tag{4.1}$$

The Bayesian approach to classifying a new instance is to assign the most probable target value given the attribute values $\langle x_1, x_2 ... x_m \rangle$ that describe this new instance [81]. The NB classifier, computes $P(X|Y)$, based on the simplifying assumption that attribute values are conditionally independent given the target value of the instance, as described by the following equation:

$$y_{NB} = argmax_{y_j \in Y} P(Y = y_j)\Pi_i P(X = x_i|Y = y_j) \tag{4.2}$$

Given y_j that denotes the target value of an instance, NB learning involves estimating the various $P(x_j)$ and $P(x_i|y_j)$ terms based on their frequencies over the training data [81].

Given our training data, we produce the parameters of a NB classifier, which we then use to predict whether a voxel is likely to become tumour by calculating the posterior probability for this voxel given the attributes of the voxel, *i.e.*, the tumour-specific information, properties of the brain, and image-based features specific to this voxel and to its neighbourhood. The classifier's output is a prediction probability that we threshold to obtain voxel labels ('tumour' or 'non-tumour').

Logistic Regression

The algorithm we used in the experiments is a multinomial Logistic Regression classifier with a ridge estimator that slightly modifies LeCessie's original algorithm [68] to handle instance weights; see [132]. But in our work, we only use binary classes; therefore, this multinomial algorithm reduces to a binomial one.

Given k classes (note in our data, we use binary classes, 'tumour' and 'non-tumour') with m attributes, the parameter matrix β to be calculated will be an $m \times (k-1)$ matrix. X is the vector of attributes describing an instance in the training data, and Y is the target value of the instance.

The probability for the first class is:

$$P(Y = y_j|X) = \frac{e^{(X \cdot \beta)}}{1 + e^{(X \cdot \beta)}} \tag{4.3}$$

The second class has probability:

$$P(Y = y_k|X) = \frac{1}{1 + e^{(X \cdot \beta)}} \tag{4.4}$$

Since we need to classify unlabeled data points into binary classes, we describe a multinomial Logistic Regression algorithm that reduces to building a binomial model (as shown in Equations 4.3 and 4.4). For further details on the implementations of this algorithm, see [132]. The output of the Logistic Regression classifier for each testing voxel is a prediction probability that we threshold to obtain the binary class label for this voxel.

Linear SVM

The linear Support Vector Machine (SVM) classifies the data by finding the linear discriminator that separates the data into two distinct classes. Linear SVM performs this classification by determining the linear separator that maximizes the margin before hitting a data point, on each side of this separator, such that the points on the two sides of the separator represent two distinct classes. Figure 4.2 shows an example classification of a linear-kernel SVM that separates the data points into two distinct classes, each class on a different side of the hyperplane. See [60] and [88] for further details and for the mathematical representation of the algorithm.

We use WEKA's implementation [132] of Platt's sequential minimal optimization algorithm for training a support vector classifier [88]. This implementation globally replaces missing values, and normalizes the attributes by default. The probability value for each classified voxel is estimated, given the decision function score, by fitting Logistic models to the margins [89, 132].

4.4.3 Training and Testing Data Folds

To perform cross-validation and patient-specific labeling, we use the leave-one-out method. Patient-specific labeling requires categorizing the data into two distinct folds: one fold used in the learning and another in the classification. The training fold includes all patients except one (*i.e.*, it includes $n-1$ patients for a dataset size n). In our dataset, the training fold contains 16 patients, some of which may have multiple studies at different time points. This training fold includes the labeled voxels from several brains. Each of these voxels is initially a non-tumour voxel adjacent to the tumour mass at the first of two consecutive studies of the patient, and labeled as 'tumour' or 'non-tumour' based on its status on a later scan. The second data fold is used for testing and consists of the voxels of the patient study (*i.e.*, the 17^{th} patient) that has been excluded in the training process. The classifier predicts

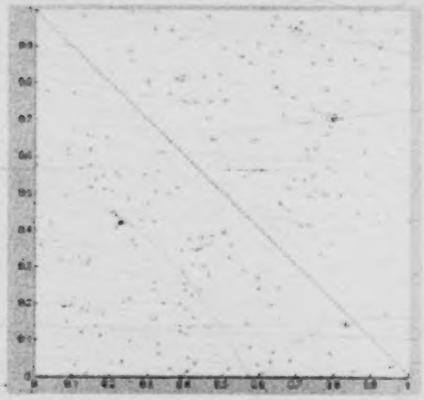

Figure 4.2: Support Vector Machine classification of a 2D linearly separable dataset
The line between the two classes is the linear discriminator found by the Support Vector Machine and the three circled data points are the support vectors.

tumour growth by labeling each testing voxels based on the model learned from the training data and the feature set used in the experiment.

Since the purpose of our work is to predict glioma growth at the patient level (using a 17-fold cross-validation), we do not perform voxel-level cross validation, *i.e.*, our model must predict tumour growth for a new patient that was not represented in the training data.

Performing patient-specific predictions (as opposed to voxel-level) allows us to observe possible factors involved in glioma diffusion and to identify special cases of tumour growth. Prediction results are converted into images to help visually inspect and explain the model's performance, and to compare the prediction of CDM with that produced by standard methods, UG and GW.

4.5 The Feature Space

We list in this section the feature set used in the prediction task. The features are grouped into five categories (as mentioned in Section 3.7). Different feature combinations are used in the experiments depending on the information content and significance of the features (see Chapter 5). This section lists all the potential features used at different stages of the experiments, described per feature category:

- Patient-related characteristics including age, pathology and history.
- Brain features describing anatomical regions.
- Tumour-specific features describing the tumour volume and growth rate.
- Features local to the voxel including image intensity information.
- Features describing the adjacent neighbourhood of the voxel.

We describe each of the feature categories, the importance of the feature for the learning task, and the challenges encountered in selecting the features. We found that to successfully model tumour growth using supervised learning, we need to answer two main questions: "What features contain the most important information that would express tumour growth? What features represent factors in glioma diffusion given the data we collected?" But answering these questions required a great deal of literature reviewing. Some of the information presented in the literature could not be used in this thesis due to lack of data and to other limitations. The data preprocessing framework and the feature extraction component have been extensively described in Chapter 3.

Throughout the experiments, regardless of the feature subset being used, we have used the same class labeling, which assigns the unlabeled voxel to one of two classes: 'tumour' or 'non-tumour' (Section 4.2). In the training stage, class labels are retrieved from the second of two consecutive patient studies.

We list the features per category; note the feature naming between parentheses as it will be used in Chapter 5 to identify the individual features involved in the different experiments.

Patient-related features

- Age: The patient's age has usually some correlation with tumour grades and survival time. More aggressive tumours tend to occur with a higher frequency in older patients, while low-grade tumours are more common in children [30, 67]. The correlation between the patient's age and the tumour grade can be observed throughout our dataset of 17 patients, as shown in Table 4.1.

- Pathology: This patient-specific feature indicates the type or grade of the glioma, and was used in preliminary experiments but was later excluded from the empirical feature space (as it can be expressed through other attributes). The glioma grade is important because it indicates the speed and invasiveness of the tumour diffusion. That is, tumours that spread faster are often high-grade tumours; in fact, the most common, and invasive tumours are grade *IV* GBM. But this information has also been expressed in terms of the growth rate and the increase in the tumour volume and border area.

- History of malignancy: This feature provides information about other malignant tumours in the patient's records, and was used in preliminary experiments only. This feature did not seem to express additional tumour growth information; in particular, our prediction is based on primary brain tumours that originate from a glial cell rather than from metastasis.

- Family history: This feature provides information about whether there is family history of glioma in the patient's records, and was used only in preliminary experiments as it did not add tumour growth information to the learning model.

- Karnofsky Performance Scale (KPS) is a numerical value that quantifies the health status of the patient. Higher KPS values are often associated with a prolonged survival time [30, 67]. KPS was initially used in early experiments but did not help

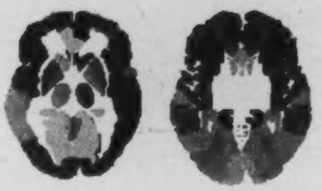

Figure 4.3: AAL and Brodmann anatomical labels
Left to right: an example image slice of Anatomic Atlas Labeling (AAL), and of Brodmann maps, both generated for 1mm^3 voxel images (different colors correspond to different anatomic labels).

express tumour diffusion information and was therefore excluded from the final feature set.

Brain anatomy features

In an early stage of the experiments, we have included in the feature set the Anatomical Automatic Labels (AAL) [122] and Brodmann maps [9, 31] to indicate the approximate anatomical or functional location of each voxel in the dataset (see Figure 4.3). We used these functional labels assuming that the tumour anatomical location might provide insight into glioma diffusion and that tumours might have different growth rates in different brain regions. But these labels were eventually excluded from the experimental feature set as they did not contribute with additional information to the model neither from a clinical aspect nor empirically. First, AAL and Brodmann labels were originally generated for templates with 1mm^3 voxels. Since we are using 8mm^3 voxel images, we generated an approximation of these brain regions. Our approximation was not sufficiently accurate and may likely have contained errors. Second, tumours spread over several anatomical regions, and in rare cases, over regions that are not labeled, which is ambiguous for the classifiers. Third, expressing these regions in terms of numerical features causes overfitting (nominal regions could not be used with SVM). And finally, there is no clinical evidence that tumours have different growth behaviours in different functional brain regions.

We also plotted image intensities obtained from three different MR modalities, T1-weighted, T1 with contrast, and T2-weighted, to examine if there are potential intensity clusters that may correspond to the different anatomical structures in the brain. The intensities were plotted for a 3D image volume, and represented by vectors $\langle T1, T1c, T2 \rangle$ where T1 is the voxel's intensity value obtained from T1-weighted image, T1c is the intensity obtained from T1-weigted image with contrast, and T2 is the intensity obtained from the T2-weighted image. The plotted intensity values included all brain voxels (*i.e.*, ventricles, skull, tumour, etc). The intensity vectors were normalized, and yielded more dense cluster-like regions that correspond to particular anatomical structures (*e.g.*, the ventricles which are dark on T1 and bright on T2) but did not show any meaningful groupings of the rest of the brain anatomy (see Figure 4.4). Clustering would have been performed with K-means algorithm [51] for several values of K that would correspond to the number of anatomical structures to be identified. But as shown on Figure 4.4, there does not exist any clear boundaries between the plotted intensities, which consequently led to discarding the option

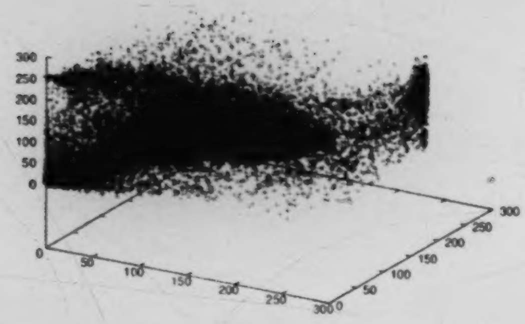

Figure 4.4: Plotting of image intensities from T1, T1-contrast and T2 modalities
The plotting was performed with intensity vectors ⟨T1,T1c,T2⟩ for a 3D image volume of 256x256 image size at the slice level.

of identifying brain anatomy regions with clustering.

Eventually, the empirical feature space does not include any features describing the anatomical brain regions since we removed the anatomic labels and the Brodmann maps. We used, however, brain tissue features at the voxel level such that a voxel may be identified as a grey matter, white matter or CSF region (see 'Features local to the voxels' below). These voxel-specific features help generally define the anatomical structures of the brain, and the regions where gliomas can grow (*e.g.*, white matter tracts) versus anatomical barriers (*e.g.*, the ventricles).

Tumour-specific features

- The 3D tumour and edema volumes (TumVol, EdemaVol) computed in terms of the number of voxels of the visible mass on the MRI scan.
- The tumour area-volume ratio (AreaVol) is the quantity $\frac{|A_t|}{|V_t|}$ where A_t is the tumour surface area and V_t is its volume.
- Tumour growth rate (GRate) computed as the volume difference observed between two consecutive patient studies for which the time interval is expressed in terms of days; *i.e.*, the growth rate is $\frac{|V_2-V_1|}{|t_2-t_1|}$ where V_2 is the volume observed at time t_2 and V_1 is the initial volume detected at time t_1. Even though GRate is an important feature that implicitly describes the tumour grade (see Table 4.1), GRate was eventually excluded from the empirical feature space as it required using multiple studies per patient in order to estimate the growth rate of the tumour. Only six of the 17 patients are represented by more than one study in the training data. Should additional studies become available for these patients, the growth rate could then be estimated from the intermediate studies of the patient (as the final study on the timeline is used only for testing).
- The edema percentage (EdemaPerc) is computed as $\frac{|100\times(V_e-V_t)|}{|V_e|}$ where V_e is the total edema volume also including tumour regions and V_t is the tumour volume. V_e

is the volume delineated from the T2-weighted image, and V_t is the volume of the enhancing tumour obtained from T1-weighted with contrast.

- The tumour volume increase (VolInc), $\frac{|V_2|}{|V_1|}$ where V_1 is the tumour volume at one time point and V_2 is the volume at a later time; in other words, this feature measures by how many times the tumour grew between the two scans.

- Volume difference (VolDiff), $V_2 - V_1$ in terms of the number of voxels

Tumour and edema volume-related features are important quantities that provide information about tumour invasiveness. Larger glioma sizes may often indicate the grade of the tumour and its capability of diffusion. Area-related features may provide information about the tumour diffusion dynamics, and about the shape of the tumour border since the existence of sharp edges or tendrils suggests further malignant invasion.

Features local to voxels adjacent to the tumour

- Probability that the voxel belongs to each of three tissue classes: white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF). This probability is obtained as a normalized intensity (*i.e.*, by dividing the image intensity corresponding to the voxel by the maximum intensity). These tissue classes are obtained from the segmentation of the patient's brain image using SPM [3, 5, 6, 104].

- Voxel membership in each of the three tissue classes (white and grey matter, and CSF). This membership is based on probabilities obtained above and is expressed as $\{0, 1\}$ based on an intensity value threshold of 50% with '1' meaning that the voxel belongs to the specified tissue class and '0' that it does not.

- Probability of the voxel's membership in each of the white matter, grey matter, and CSF, (Ch2GM, Ch2WM, Ch2CSF) obtained by segmenting the CH2 template [54] into the tissue classes using SPM.

- Edema expressed as $\{0, 1\}$ to indicate whether this voxel is located in an edema region or not.

- Euclidean distance (EuclDist) is the minimum distance from the local voxel to the tumour border computed as $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$

- Coordinates x,y,z of the local voxel in the brain image

- T1 and T2 intensity values (T1Int, T2Int)

- The intensity difference between T1-contrast and T1: tumour regions that enhance on T1-weighted with contrast images will yield a non-zero intensity difference.

- Intensity value (Ch2Int) obtained from CH2 template [54] to which we registered the patients' scans.

- Distance-area (DistArea), computed as the $\frac{Edist}{A_t}$, where $Edist$ is the Euclidean Distance between the current voxel and the border of the original tumour, and A_t is the surface area of this tumour.

- Distance-volume ratio (DistVol) is $\frac{Edist}{V_t}$ where $Edist$ is the Euclidean Distance between the current voxel and the border of the original tumour, and V_t is the tumour volume.

- Tissue-map membership $\{0, 1\}$ where '1' indicates that the voxel is located in a tissue region where tumours grow, and '0' indicates that the voxel is located in a bone, ventricle, or membrane region where tumours cannot grow. The tissue map was generated from segmenting the CH2 template [54] and the average T1 template [79] with SPM [104], and overlaying the grey and white matter obtained from both segmented templates.

Neighbourhood features

In 3D, each voxel has 6 immediate neighbours, that is, 2 adjacent voxels from each dimension. The following features are computed for each of the 6 neighbours of each training voxel in the dataset:

- Probability for each of the 6 neighbour-voxels to fall into an edema region (EdemaNei), computed from the normalized intensities of the edema volume after registration and spatial interpolation (*i.e.*, a total of 6 features where a neighbour outside of the brain boundary has zero probability value).

- T2 and T1-contrast scaled intensity values (T2IntNei, T1cIntNei) obtained from the patient's image, and computed for each of the 6 neighbour-voxels.

- Membership probabilities for each of the voxel's neighbours in each of the three tissue classes: grey matter, white matter, and CSF, segmented both from the patient's scan (GmNei, WmNei, CsfNei) and from the CH2 template (Ch2GmNei, Ch2WmNei, Ch2CsfNei) with SPM providing 36 probability values in total.

4.6 Diffusion Models

We describe in this section the classification-based diffusion model CDM we developed for the purpose of this thesis as well as two standard models that we implemented to compare our results. One of these standard methods describes the uniform 2cm boundary conventionally used in specifying the treatment volume [48, 56, 126]. We implemented this uniform model to determine if tumours grow spherically. The second standard model assumes that the diffusion rate of malignant cells is different depending on whether the tumour is spreading in the heterogeneous grey matter or white matter tissue in the brain [110]. All three implementations are designed to allow the tumour to grow on the voxel level, but not outside of the brain border.

Patient Id	Patient Age	Tumour pathology (Grade)	GRate ($\frac{cu\ voxels}{days}$)
6	33.1	Astrocytoma *II* (likely progressed into GBM)	210.7
8	38.4	Anaplastic Astrocytoma *III*	14.7
9	54.9	Glioblastoma Multiforme *IV*	80.3
25	60.1	Glioblastoma Multiforme *IV*	55.6
37	59.9	Anaplastic Astrocytoma *III*	17.6
44	13.4	Astrocytoma *I*	5.2
51	50.0	*Unspecified* (likely GBM)	110.7
69	37.7	Mixed Oligo-Astrocytoma *II*	118.6
83	54.4	Glioblastoma Multiforme *IV*	31.7
93	77.6	Glioblastoma Multiforme *IV*	102.5
105	52.9	Glioblastoma Multiforme *IV*	28.5
106	70.4	Glioblastoma Multiforme *IV*	25.8
108	58.0	Glioblastoma Multiforme *IV*	58.0
147	41.9	Glioblastoma Multiforme *IV*	24.4
160	72.0	Glioblastoma Multiforme *IV*	51.4
165	41.5	Glioblastoma Multiforme *IV*	133.6
169	65.9	Glioblastoma Multiforme *IV*	52.8

Table 4.1: Tumour grade and growth rate information for each patient in the dataset
This table provides the patient's age, the tumour grade information (or histology) and the growth rate (GRate), computed using the first and final studies on the timeline, for each of the 17 patients that represent the training data. Note the correlation between the patient's age and the tumour grade (*i.e.*, low-grade gliomas are more common in the younger population and in children while grade *IV* gliomas occur more frequently in older patients [2, 30]). Also, GBM (grade *IV*) tumours are the most common in glioma patients [30, 130], and represent therefore $\frac{2}{3}$ of our training data.

```
Diffusion( VoxelLabel:  VL; GeneralInfo:  e; int:  s)
    % VL[i, j, k]=1 if position $\langle i, j, k\rangle$ is a tumour
    % Initially VL corresponds to current tumour
    % When algorithm terminates, VL will correspond to tumour containing "s" additional voxels
    total_count := 0
    Do forever:
        Compute $N := \left\{ \langle x, y, z\rangle \;\middle|\; \mathrm{VL}[x, y, z] = 0 \;\&\; \begin{pmatrix} \mathrm{VL}[x+1, y, z] &=& 1 & \vee \\ \mathrm{VL}[x-1, y, z] &=& 1 & \vee \\ \vdots \\ \mathrm{VL}[x, y, z-1] &=& 1 \end{pmatrix} \right\}$
        For each location $v_i \in N$
            Determine if $v_i$ becomes a tumour                    (*)
            If so,
                Set VL[$v_i$] := 1
                total_count++;
                If (total_count == s), return
```

Figure 4.5: Generic Diffusion Model

In general, a diffusion model takes as input the current 'voxel label' (VL), which labels a voxel as '1' if it is currently a tumour and '0' otherwise (see Figure 3.7), and general information $\mathbf{e} = \mathbf{e}_{Patient} \cup \mathbf{e}_{Tumour} \cup \{\mathbf{e}_i\}_i$ about the patient $\mathbf{e}_{Patient}$, the tumour $\mathbf{e}_{Tumour}$ and the individual voxels $\{\mathbf{e}_i\}_i$. The third input s tells the diffusion model how far to grow – *i.e.*, s is the number of additional voxels to include. The output is the prediction of the next s additional voxels that will be incorporated into the tumour (*e.g.*, imagine the tumour is currently 1000 voxels and the doctor needs to know where the tumour will be, when it is 20% larger – *i.e.*, when it is 1200 voxels. We would set $s = 200$). See Figure 4.5.

A diffusion model is first initialized from the original tumour at each stage. The diffusion model identifies the set of voxels N just outside the border of the initial tumour; see Figure 4.5. In the following image

v1	v2	v6	v7	v5
X	X	v3	v4	X
X	X	X	X	X

(4.5)

(where each **X** cell is currently a tumour cell), N would consist of the voxels labeled v1 through v5, but not v6 nor v7 (as we are not considering diagonal neighbours). In the 3D case, each voxel will have 6 neighbours.

The diffusion model then iterates through these candidate voxels, v_i . If it decides that one has become a tumour, it then updates VL (which implicitly updates the tumour/healthy border, and increments the total number of 'transformed voxels'). Note we do not consider edema, nor any other labels.

After processing all of these neighbours (in parallel), it will then continue transforming the neighbours of this newly enlarged boundary. If a voxel is not transformed on one iteration, it remains eligible to be transformed on the next iteration. When the number of transformed voxels matches the total s, the algorithm terminates, returning the updated VL assignment.

The various diffusion models differ only in how they determine if v_i has become tumour – line ($\star$) of Figure 4.5. The *uniform growth* model, UG, simply includes every 'legal' voxel it finds (where a voxel is legal if it is part of the brain, as opposed to skull, eye, etc). The *tissue-based* model, GW, assumes the growth rate for white matter is 5 times faster than for grey matter [110], and 10 times faster than other brain tissue. In the GW model, whenever v_i is white matter, it is immediately included. If v_i is grey matter (other tissue), its count is incremented by 0.2 (resp., 0.1) on each iteration, and the voxel is included when its count reaches 1.0 . GW does not allow diffusion into the skull. The $\mathbf{e}_i$ part of the `GeneralInfo` e specifies the tissue type of each v_i voxel, as computed by SPM [104].

CDM uses a probabilistic classifier to compute the probability q_i that one tumour neighbour of this voxel v_i will transform this voxel to become tumourous, $q_i = P_\Theta(\,\ell(v_i) = \text{tumour} \,|\, \mathbf{e}_{Patient}, \mathbf{e}_{tumour}, \mathbf{e}_i\,)$ where Θ represents the parameters of the model (*i.e.*, the classifier's CP table). Some voxels can have more than *one* such tumour-neighbours; *e.g.*, in (4.5), the voxels `v1`, `v2` and `v5` each has a single tumour-neighbour, while `v3` and `v4` each has 2. Each tumour-neighbour of the voxel v_i has a q_i chance to transform this v_i; hence if there are k such neighbours, and each acts independently, the probability that v_i will be transformed on this iteration is $p_i = 1 - (1 - q_i)^k$. CDM will then stochastically transform this voxel to be a tumour with probability p_i . CDM performs these computations in parallel – hence even if `v3` is transformed, `v4` still has only 2 tumour-neighbours (on this iteration). CDM learns q_i by training on the learning data folds. One could argue that at training time, we could be learning p_i (rather than q_i), we are exploring this issue as a future research direction.

Another alternative to stochastically simulating tumour growth is to threshold the classification probability values where values above the threshold would be associated with a positive (*i.e.*, tumour) voxel, and values below the threshold are negative voxels. A threshold of 65% was used, that is by satisfying the condition $p_i > 0.65$ and yielded slightly higher results than the stochastic simulations. This is because low-probability voxel candidates were excluded from the tumour class. Such candidates can often be misclassified.

Other thresholds of 60% and 70% were used as well but a lower threshold tends to include more false positives and higher thresholds yielded similar results to the 65% threshold. Also, higher thresholds make the diffusion algorithm more computationally costly as the number of iterations increase because with higher thresholds of 70% or above, a fewer number of voxels per iteration meets this threshold, and so more iterations are required to reach the number of additional voxels s to include.

A 10% diffusion rate was incorporated into CDM to help model the mass effect of tumour growth on deformed CSF regions (*e.g.*, the ventricles and sulci); see Figure 4.6. This 10% rate was selected based on the percentage overlap between the CH2 template [54] (a normal brain with no tumour) and each of the patients' scans where the tumour has induced pressure on the ventricles. The average overlap percentage obtained for several patients was approximately 10% (Figure 4.6).

We empirically evaluated the three models, UG, GW and CDM, over a set of 17 patients. For each patient, we had two sets of axial scans R_1 and R_2 taken at different times, each with known tumour regions. Let s_i refer to the size of the tumour in scan R_i. We then input to each model the R_1, and predicted the next $s = s_2 - s_1$ voxels that would be transformed. We could then compare the predicted voxels, with the truth – *i.e.*, the tumour region of the

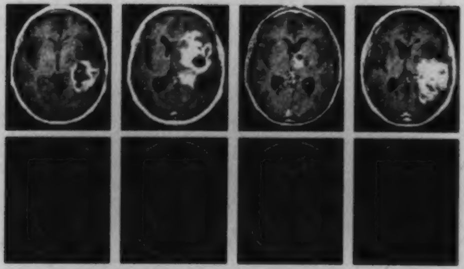

Figure 4.6: Tumour-induced pressure on the ventricles

Top: image slices for four different patients showing tumour-induced pressure on the ventricles. Bottom: the corresponding slices from a template of a normal brain. Note: the induced pressure in the patient's scans is represented by the tumour overlap with the ventricles in the template. The average overlap for a sample of nine patients was approximately 10%.

second scan, R_2.

To measure the quality of each model, let "nt" be a set of new tumour cells for a patient (*i.e.*, this is the 'truth') and "pt_χ" be the cells that model χ predicts will be transformed. The standard measures we use are 'precision' of χ (on this patient) is $\frac{|nt \cap pt_\chi|}{|pt_\chi|}$ and 'recall' is $\frac{|nt \cap pt_\chi|}{|nt|}$. In our case, as our diffusion models stop when $|pt_\chi| = |nt| = s$, the precision and recall values will be the same (see Section 5.3) as opposed to values directly obtained as classifiers' results where there is usually a significant tradeoff between the precision and the recall values. Also, the diffusion models do not allow gaps in the simulation; that is, each of the model iterates around the initial tumour border producing a contiguous, encapsulated tumour mass that has the same volume as observed on the second of two consecutive scans of the patient (*i.e.*, the 'truth' volume for that patient).

While UG and GW are completely specified already, CDM must first be trained. We use a 'patient level' cross-validation procedure. That is, we trained the classifier (*e.g.*, Naïve Bayes [36]) on $n-1$ patient studies, then tested on the n^{th} study. Each training instance corresponded to a single voxel v_i outside the initial tumour in the first scan R_1, with features $\mathbf{e}_{Patient}$, $\mathbf{e}_{tumour}$, and $\mathbf{e}_i$, where the label was '1' if this voxel was in the tumour in R_2, and '0' otherwise. Training is at the voxel level, and is not based on diffusion, but the tumour volume considered in the training corresponds to the ideal scenario of diffusion.

All diffusion models allowed the tumour to grow to the volume observed on the second of two consecutive scans of the same patient.

While virtual simulations often assumed that tumours are self-limited systems that grow uniformly in all directions, as described in Chapter 2, our UG model was tested on clinical data to determine if tumour growth behaviour is generally non-uniform. UG restricts virtual tumour growth outside of the brain border and into the skull, and tends to preserve the shape of the initial tumour producing a larger mass that has the exact same shape of the initial tumour detected on the first scan. In GW, we generated the grey and white matter, and CSF tissue classes by segmenting the average T1 template with SPM [104], which gave more contiguous tissue classes in comparison with those obtained by segmenting the CH2 template [54] or the patients' brain image.

4.7 Summary

Modeling glioma diffusion is not a simple task. In spite of the tremendous research efforts over the last few decades, scientists are still struggling to understand the factors involved in tumour growth and its capability of invasion. But these research efforts have made many contributions toward the improvement of treatment methods and planning. In particular, the use of software algorithms and tools has helped medical doctors view the patient's brain in 3D and target the tumour volume more properly with computer-assisted radiation planning.

In this chapter, we described how we model glioma growth from a machine learning perspective based on factors which, according to state-of-the-art literature, are believed to play an important role in glioma diffusion and invasion. The task of modeling glioma growth involves four important components. First, the type of classifier involved in developing the proposed model has a significant impact on the model's results; this is because different machine learning algorithms are sensitive to the features depending on how the decision function computes the prediction for a particular data instance. Second, the feature space represents an important part of this thesis as it required investing a few months into research and experiments in order to find out what factors may have a role in glioma diffusion and how these factors can be expressed in terms of training features to help describe tumour growth more accurately. The prediction is ultimately based on features specific to the patient and the tumour, and attributes of each candidate local voxel and the neighbours of the candidate voxels. Third, the processing of the empirical data has also been a significant component of the work in this thesis. In order to extract the classification features for each patient from the clinical dataset collected from the CCI, the MRI scans required several steps of preprocessing and denoising as described in Chapter 3. Finally, we developed a diffusion algorithm that uses the probabilities predicted by the classifier to simulate tumour growth. The algorithm is initialized from the patient's tumour and allows the tumour to grow to a specified target volume (for testing, we set this target volume to the volume detected on a later scan) by iteratively adding to the tumour border the eligible voxels that meet the probability condition as shown in Section 4.6.

Chapter 5

Results

This chapter presents the experiments, the results, and challenges involved in assessing the performance of our implementation of glioma diffusion modeling. Section 5.1 describes the performance measures used in evaluating the implementation and the challenges of properly assessing the model's performance as compared to existing approaches. Section 5.2 describes the feature selection method and illustrates how we selected the different subsets of features involved in the experiments and the significance of these features. Section 5.3 presents the experiments followed by an in-depth analysis of the results for each experiment, including a description of average and best case results versus special scenarios of tumour diffusion. This section also evaluates the dataset and the training features in light of the empirical results, and discusses issues related to the performance measures.

5.1 Performance Evaluation of CDM

In this section, we first describe issues involved in evaluating CDM's performance and we then discuss the evaluation measures used in quantifying the model's accuracy.

Assessing the performance of our implementation of glioma diffusion modeling has been a challenging task due to several factors and data-related issues:

- Lack of implementation of standard methods
- Unavailability of standard datasets
- Inconsistency of manual tumour segmentation
- Method of generating the training data
- Error margin in evaluating performance

We further describe each of the above listed issues in more detail. We also explain whether we resolved the issue or if it remains an open question for future research.

First, there are no available implementations of existing methods that we could use to compare the performance of CDM to that obtained by standard methods. For this reason, we implemented two of the standard models: uniform radial diffusion, UG, and tumour diffusion based on motility in grey and white matter, GW. The UG model is conventionally used in radiotherapy and implicitly assumes that the tumour grows in all directions equally (our implementation does not allow the tumour to grow outside of the brain border). The GW growth model is suggested by Swanson et al. [110] and assumes faster diffusion in white versus grey matter with $5 : 1$ ratio (see Section 4.6). In all of the experiments, we compare the accuracy of the three models to clinical data for each patient. That is, we compare the tumour growth predicted by each of UG, GW, and CDM to real tumour growth observed on the patient's brain scans.

Second, to obtain a consistent assessment of models' performance and comparison, it is required to evaluate the various models on a standard dataset. However, there is no standard dataset commonly available for this type of research. This is mainly due to confidentiality of patient scans and data privacy policies set by governments and health institutions. As a consequence for the lack of a standard dataset, different models are being evaluated on the different datasets that are available to these models.

In addition, patient treatment (often surgery and radiation) imposes a limiting factor on how the data could be used. Because of treatment, it is possible to observe tumours shrink then recur in regions adjacent to the original tumour. It is also possible to observe a decrease or elimination in fluids (that can be erroneously detected as edema on T2-weighted images while they are actually post-surgical bleeding). This effect is particularly more pronounced in brain scans taken after surgery where scans that immediately follow surgery will show high-signal regions around the original tumour because of bleeding. These high-signal regions will disappear over time in later scans. Ideally, it is most useful to collect data for untreated patients but such data is very rare and often unavailable. In our dataset, every patient has undergone a combination of one or more treatments including surgery and radiation. While it is not feasible to maintain a perfectly consistent dataset where tumours do not decrease in volume or recur, after treatment, in regions adjacent to the original tumour mass, we have put together a small but fairly consistent dataset. We have also excluded subjects where tumour shrinkage was observed or where surgical cavities introduced additional noise in the image.

Third, we obtained the delineated tumour and edema volumes through manual labeling of the brain images by drawing the contour of the tumour volume on the images before performing data processing. Tumour labels were drawn with the help of a tool that was locally implemented and that allowed users to view the different image modalities simultaneously. Segmented images where tumour and edema volumes where manually delineated were verified by radiologists. These manual labels are generally more accurate than automated segmentations of tumour volumes but are likely less consistent than automated labeling because of the error margin at the tumour boundary and because of human subjective interpretation of abnormal regions on brain scans. In spite of these issues, several existing tumour growth models use manually delineated data as it is more accurate than automatically segmented volumes (Chapter 2). In our work, since we use learning and classification, our model results will depend on what the model learns as 'tumour', and which in turn, depends on the labeling decisions made by human experts.

Fourth, deciding how to generate the training data is a significant factor in evaluating model performance as the model relies on supervised learning. Generally, tumour voxels represent only a small percentage of the total number of voxels in a brain image, and therefore, positive data instances will not be well represented in the training data if the entire brain image was used in the training. Using the entire brain image would result in very large datasets that would make the training task rather computationally costly. For these reasons, sub-sampling is required in order to obtain a more balanced training set. In our case, we have trained classifiers on voxel regions adjacent to the original tumour at the first of two consecutive time points (but that exclude the tumour itself) up to a 2-voxel border around the tumour at the second time point (*i.e.*, when the tumour has reached a larger volume); see Chapter 4. This set of voxels used in training represents the voxels considered by a diffusion model, using a *perfect* classifier. This selection of training data allows a stronger representation of the positive (tumour) voxels in the learning phase.

In addition to the selection of training voxels, it is worth noting that using different subsets of the feature space may help improve model performance depending on the classifier used in the learning and the prediction (Sections 5.2 and 5.3). This is because some tumour diffusion scenarios require a higher degree of generalization for the learned model which is determined by both the classifier used and the feature set, given the training data.

Finally, since the learning and classification are performed at the voxel level, we also evaluate model performance at the voxel level using the precision and recall measures [12]. This voxel-level assessment will include an error margin at the tumour boundary due to segmentation error and to misclassifications of the tumour-healthy border regions.

We define the precision $P = \frac{|nt \cap pt_\chi|}{|pt_\chi|}$ and the recall $R = \frac{|nt \cap pt_\chi|}{|nt|}$ where "nt" represents the cells that have become tumour for the patient (*i.e.*, the truth) and "pt_χ" is the set of tumour cells predicted by the model. Because the diffusion model generally stops when $|pt_\chi| = |nt| = s$, then $P = R$, except when the model terminates early as the tumour grows into a significantly large volume, or because of segmentation errors at the tumour boundary, or in cases of recurrence (*i.e.*, the tumour shrinks but later grows in a region adjacent to its original location); see Section 5.3. Achieving equal precision and recall scores compensates for the tradeoff between these two measures when directly obtained as classifiers' results [12].

5.2 Feature Selection

In preliminary experiments, we used a feature set, IFS, listed in Table 5.6, that combined four of the five feature categories described in Section 4.5. IFS included attributes specific to the patient and labels for brain anatomy regions and tissue types. It also included information about the volumes and surface areas of the tumour and the edema, the distance between normal voxels and the tumour border, and intensity-based features local to the normal voxels adjacent to the tumour. This feature set did not incorporate neighbourhood attributes for the training voxels – *i.e.*, the normal voxels adjacent to the tumour (see Table 5.6). This initial feature set was applied to a 5-patient dataset with patient-fold cross-validation. Results showed that some of the features involved in preliminary experiments did not accurately express glioma diffusion in most scenarios (see Table 5.1). In particular, features such as patient history and Karnofsky Performance Scale (KPS), anatomical labels, voxel coordinates, and raw tumour volumes computed as the number of tumour voxels per patient, did not help our classification-based model find diffusion patterns in regions ad-

Patient Id (Pathology)	CDM	UG	GW
9 (GBM)	0.753/0.444	0.617/0.364	0.872/0.514
37 (AA)	0.381	0.330	0.421
69 (MOA)	0.280	0.288	0.334
85 (AA)	0.274	0.259	0.239
160 (GBM)	0.744	0.628	0.678

Table 5.1: Precision and Recall obtained in initial tests for the three models using the IFS feature set (listed in Table 5.6) with 5-fold 'patient-level' cross-validation

(Note: Due to unusually large tumour size in patient #9, CDM predicted only 44% of the tumour yielding different precision 0.753 and recall 0.444 . To allow comparison, we make UG and GW predict the same volume increase that CDM predicted. In this experiment, we used Naïve Bayes classifier in CDM. Pathology was used as a feature to indicate the tumour grade – Glioblastoma Multiforme (GBM), Anaplastic Astrocytoma (AA), and Mixed Oligo-Astrocytoma (MOA).

jacent to the tumour. This conclusion was based on training machine learning algorithms with various combinations of feature groups derived from the feature space listed in Section 4.5. Results did not significantly change if patient history, KPS and pathology, voxel coordinates and tumour volumes were excluded from the training features. That is, these features did not add sufficient information to the learned model, and in particular anatomical labels reduced accuracy in some scenarios, likely due to errors in approximating brain anatomical regions. For this reason, we excluded these attributes from the empirical feature space, and we then incorporated neighbourhood features that describe the voxels adjacent to each training voxel; training data represents the set difference between a pair of tumour volumes detected on two scans of the patient at different times; see Figures 1.1 and 4.1.

Having incorporated the voxel neighbourhood features listed in Section 4.5, we performed several experiments with different classifiers and with three combinations of feature groups: FS_0, FS_1, and FS_2. Based on preliminary results and the information content of the learning features, we believe that these neighbourhood features help express more accurately glioma diffusion patterns, given our training set of 17 patients. Since features have a direct role in the results obtained through supervised learning and classification, the same features may yield different results when applied to different classifiers depending on how the learning algorithm interprets the information content in the features given the training voxels. In addition, some tumour growth scenarios may require a higher-level of generalization (*i.e.*, may require features that help more properly generalize tumour growth patterns, *e.g.*, without making assumptions about the tumour location, which is specified in terms of voxel coordinates and features specific to the brain tissue). It is therefore not beneficial to apply the entire feature space since a particular combination of features may help more properly predict some tumour growth scenarios but may not as accurately predict specific scenarios, *e.g.*, where tumours induce pressure on the ventricles and deform them (see Figures 5.5, 5.7 and 5.9). In such cases, it is more useful to actually exclude voxel coordinates and tissue-specific features. To obtain various combinations of feature groups, it is necessary to first quantify the information content in the feature space, to rank the features in order of significance, and to group them based on feature types that can be used together.

We can then use these feature groups in various experiments and compare the significance of the attributes according to empirical results.

For this purpose, we computed the information gain for the feature space using the attribute selection implementation in WEKA [132], and obtained information gain scores for each attribute in the FS_0 feature set (see Table 5.6 at end of chapter) given the entire training set of 17 patients. We first describe how information gain scores are computed. Let Y be a target value among data instances where X, an attribute, has value x. The information gain, $IG(Y|X) = H(Y) - H(Y|X)$, is computed using the following equation:

$$H(Y = y_j|X = x_i) = -\Sigma_{x_i}\Sigma_{y_j} P(Y = y_j, X = x_i) \log \frac{P(Y=y_j,X=x_i)}{P(Y=y_j)P(X=x_i)} .$$

We list the results of the attribute selection in Table 5.2 (at end of chapter). Score values slightly changed when measuring the information gain on a large subset of the training data, *e.g.*, when computing the information gain for a set of 16 patients. Based on general observation of the information gain scores, we rank the feature groups, listed in descending order of significance (*i.e.*, the higher the number, the lower the significance), according to the scores in Table 5.2, as follows. (Note the description of each feature can be found in Section 4.5.)

1. Distance-area (DistArea) – distance between the local voxel and the tumour border divided by the tumour surface area

2. Edema features (Edema and EdemaNei) – the probability of edema in the 6 adjacent neighbours to the training voxel, and the edema $\{0,1\}$ for the local voxel

3. T1-weighted image intensity for the training voxel (T1Int) where abnormalities and edema may appear as dark regions

4. The voxel-level tissue map $\{0,1\}$ differentiating between regions where tumour can grow such as grey and white matter or where tumours do not grow such as bone or eyes (TissueMap)

5. Euclidean distance (EuclDist) between non-tumour voxels and the tumour border

6. Intensities obtained from T2-weighted images for the training voxel (T2Int) and its 6 adjacent neighbours (the 6 T2IntNei features)

7. Intensity value (Ch2Int) for the training voxel obtained from the Colin Holmes, CH2, template [54]

8. Tumour volume increase (VolInc)

9. Grey matter probability for the local voxel and for its adjacent neighbours obtained from the CH2 template (Ch2GM and the 6 Ch2GmNei features)

10. Intensity values obtained from T1-contrast images (T1cInt and the 6 T1cIntNei)

11. Voxel coordinates in 3D

12. Edema Percentage (EdemaPerc) and area-volume ratio (AreaVol)

13. Patient's age

14. Grey matter probabilities (GM and GmNei) obtained from the segmented brain image of the patient

15. White matter probability for the local voxel and for its adjacent neighbours obtained from the CH2 template (Ch2WM and the 6 Ch2WmNei features)

16. White matter probabilities (WM and WmNei) obtained from the patient's scan

17. CSF probabilities from CH2 (Ch2CSF and Ch2CsfNei) and probabilities specific to the patient's segmented image volume (CSF and CsfNei)

But in order to divide the features into more meaningful groups to be combined in the experiments, given the difference in information gain scores is minimal between the features (see Table 5.2), we use a "high-level" grouping of the same features ranked above. This grouping is based on the above ranking and on the information content (Table 5.2). The reason for grouping the features is to be able to identify the attributes that can be used together to express particular glioma growth information. For example, tumour-specific features are grouped together, and can often be used as one feature entity in the learning task. Another feature entity is the tissue map and voxel coordinates that may be more informative when used together. Similarly, edema attributes are used as a single feature entity, and intensity-based features are also more expressive when used together. The grey and white matter features are combined into a single group of brain tissue where tumours grow, while CSF attributes are separated into a different group since tumours cannot grow into CSF regions. The attributes (grey and white matter, and CSF) obtained by segmenting the CH2 template [54] – an image of a normal brain – are grouped separately from those obtained by segmenting the patient's image since segmenting the patient's image often yields less accurate attributes due to the presence of tumours in the images.

We list the feature groups in descending order from most to least important depending on whether a feature group should be excluded from the empirical attribute set:

- Tumour volume, area and distance information (DistArea, EuclDist, AreaVol, EdemaPerc and VolInc)
- Edema features (Edema and EdemaNei) – contain the highest information gain scores in the entire feature set
- Intensities obtained from T2-weighted images (T2Int and T2IntNei), also capable of detecting edema, and T1-weighted image (T1Int, T1cInt, and T1cIntNei) where we use the difference between T1c and T1 intensities for the local voxel rather than the raw T1cInt.
- The patient's age
- Voxel-level map (TissueMap) and the x,y,z coordinates of the local voxel

- Segmented grey and white matter of the CH2 template (Ch2GM, Ch2WM, Ch2GmNei, and Ch2WmNei)
- Segmented matter specific to the patient (GM, WM, GmNei, WmNei)
- CSF features from CH2 and from the patient (Ch2CSF, Ch2CsfNei, CSF and CsfNei) are the least informative features in the entire feature space

Using this "high-level" grouping of the FS_0 feature space (which contains 75 features), we have put together two feature subsets, listed in Table 5.6 (at end of chapter), which we use in the experiments described in Section 5.3. The first subset, FS_1, uses volume-related attributes, edema information, intensity-based features and the patient's age (*i.e.*, the first four items in the high-level feature grouping above). We excluded from FS_1 all attributes specific to the brain tissue since these attributes have relatively lower IG scores (see above feature grouping and Table 5.2). The second subset, FS_2, uses the same features in FS_1 in addition to voxel coordinates, tissue map, and features obtained from CH2 template (Ch2Int, Ch2GM, Ch2WM, Ch2GmNei, and Ch2WmNei). In FS_2, we exclude grey and white matter features obtained from patients' images as well as all CSF features (*i.e.*, features with lowest IG scores as shown in Table 5.2). Patient-specific tissue segmentation (grey and white matter, and CSF) likely contains errors because SPM [104] does not account for the presence of tumours when it segments the brain images of the patients into the three tissue classes. It is therefore more useful to obtain tissue features from a segmentation of the CH2 template, which is an image of a normal brain. We completely excluded CSF features from both FS_1 and FS_2 since these features have the lowest IG scores among all training features. These low IG scores may be attributed to tissue segmentation errors with SPM [104] due to the presence of the tumour in the patients' scans. Also, while tumours cannot grow into CSF regions (*e.g.*, the ventricles and the sulci), large tumours often induce pressure on these anatomic structures and deform them, producing an effect that there is tumour at CSF voxel regions. This effect can cause confusion during the classification task.

By excluding from FS_1 patient-specific and template-specific information such as voxel coordinates, tissue map labels, grey and white matter, and CSF attributes, we allow the model to further generalize tumour diffusion patterns. We expect this feature set, FS_1, to help more accurately predict tumour diffusion scenarios that require a higher degree of generalization. But we also expect that accuracy may slightly decrease for subjects that rely on more specific training information. In FS_2, we have included patient and template-specific information with the exception of patient-specific grey and white matter attributes and CSF, since these features have the lowest information gain scores and may not therefore properly represent brain tissue classes due to SPM's segmentation errors. Since FS_2 does not provide the same level of generalization as FS_1, it is expected that results for experiments performed with FS_2 will be more similar to the results performed on the entire feature set, FS_0.

5.3 Experiments and Results Analysis

Given the nature of the problem, we perform only patient-specific testing; *i.e.*, we train on several patient studies then we test on a study that was not represented in the training data.

This allows us to perform both inter-patient and intra-patient testing at the 3D level.

Our training data is limited by several factors (discussed in Section 4.3). Because of these data limitations, we selected only a subset of the available data for the experiments. The empirical subset represents only 17 of 200 available patients. Six of the 17 training patients are represented by more than one study in the data, which allows intra-patient testing in a way that training is performed on 16 patients in addition to one or more studies of the testing patient (the 17^{th} patient). Then, we test on the held-out study of the testing subject. In inter-patient experiments, the unlabeled patient is not represented in the training (this is the case for subjects that have only one study). Experiments are performed with cross-validation in a way that we hold out one patient study in each round of training (from 16 patients) and we classify the unlabeled held-out subject. The training set used in building the learning model usually consists of approximately $\frac{1}{2}$ million voxels obtained from 16 training patients. Experiments performed with SVM required further sub-sampling of the training data to help improve the efficiency of the training task (see Section 4.4). Training data is obtained from regions adjacent to the tumour at the first of two consecutive scans (see Section 4.2, and Figures 1.1 and 4.1). The testing set consists of the entire brain image (excluding the tumour at the first time point) for the unlabeled patient.

For the purpose of the prediction task, we use three different classifiers, available through WEKA's implementations [60, 68, 88, 132]:

- Naïve Bayes (NB)
- Logistic Regression (LGT)
- Support Vector Machines (SVM) with linear kernel

The choice of classifiers is based on two factors: computational efficiency (*i.e.*, the amount of time the learning algorithm spends on training) versus the accuracy of the classification (*i.e.*, how correct is the classifier in predicting tumour voxels for the unlabeled subject). In initial experiments, we also used Decisions Trees (C4.5) and Multilayer Perceptron (MLP) available through WEKA's implementations [132]. In initial experiments performed with C4.5 and MLP, results were similar to those obtained with other classifiers but training was more computationally costly. Therefore, we only report diffusion model results based on training and classification with NB, LGT, and SVM using cross-validation on the 17 patients.

Once we have obtained the labels predicted by the classifier (*i.e.*, 'tumour' or 'non-tumour') for each of the voxels of the testing patient, we run a diffusion algorithm on this patient. The algorithm decides whether a voxel will become 'tumour' based on the classifier's labels and the probability of the label at this voxel. The diffusion algorithm is initialized from the tumour at the first of two consecutive time points for this patient, and iterates around the active border of the tumour until it reaches the target tumour volume (*i.e.*, upon adding k voxels to the initial tumour volume). For testing, we allow the model to grow the tumour to the volume observed at the next time point. The algorithm iterates until it reaches the volume observed on the next scan of the patient or until it is no longer able to add more tumour voxels to the predicted tumour volume (see Section 4.6). The latter situation is more likely to occur if the classifier's probabilities are very low (*i.e.*, if the classifier was not able to decide whether a voxel will become tumour resulting in a large number of false negatives). As the diffusion algorithm continues to iterate and as more voxels become 'tumour', a low-probability voxel v may also become 'tumour' if one or more of this voxel's neighbours have already become 'tumour', *i.e.*, if v is situated in a concave region of the tumour border; see Diagram 4.5 and Figure 4.5. But when the number of false negatives is large, CDM may iterate for very long periods of time (in terms of days) before it reaches the volume at the next time point. In these cases, the diffusion algorithm is able to add only a very small number of voxels to the tumour in a single iteration, requiring a larger number of iterations in order to reach the specified volume. This issue is even more prominent if the real tumour has grown into a significantly large mass in comparison with its initial volume at the first time point. In cases like these, it may not be possible to allow the diffusion model to iterate for lengthy periods of time (in terms of days), but the algorithm will be terminated early though it may have predicted only a fraction of the specified tumour volume. This is the case of the prediction output obtained with Naïve Bayes classification (see Tables 5.4 and 5.5).

The results of our CDM model are compared to those of UG and GW. These two models do not involve as many parameters and are not as sophisticated as CDM. They consist basically of initializing a diffusion algorithm from the tumour at the first time point. The algorithm iterates around the tumour border adding 'eligible' voxels to the tumour mass, until the tumour reaches the volume observed at the next time point (see Section 4.6).

So far in this chapter, we identified the performance evaluation measures, and we described the training data, the features, and the classifiers used. In the remainder of this chapter, we describe the experiments and the results. In Section 5.3.2, we analyze the average and best case scenarios where CDM performs more accurately than UG and GW. In Section 5.3.3, we discuss special scenarios where CDM performs the same as the other models. We then evaluate the training dataset and the features in Sections 5.3.4 and 5.3.5 respectively, and we discuss the fairness of the performance measures in Section 5.3.6. We conclude the chapter with remarks about the problem of glioma growth prediction and the performance of our CDM system.

5.3.1 Experiments and Results Comparison

In this section, we present three rounds of experiments performed with training on different feature combinations, FS_0, FS_1, and FS_2 (see Table 5.6) which we derived from the feature space listed in Section 4.5, based on the feature selection method described in Section 5.2. Experiments have been performed with Naïve Bayes (NB), Logistic Regression (LGT), and linear-kernel Support Vector Machines (SVM). In each experiment, we compare our CDM model results with results obtained with UG (uniform diffusion) and GW (tumour growth based on diffusion motility in grey and white matter).

We report results obtained for CDM with cross-validation for the 17 patients. The dataset consists of various tumour grades ranging from low-grade astrocytomas to grade *IV* glioblastomas (GBM). Note that we excluded the *pathology* attribute from the empirical feature space (Section 4.5). Though the tumour grade helps indicate the degree and speed of invasion, the extent of invasion can be expressed more properly in terms of the growth rate and the tumour volume increase (*i.e.*, feature attributes GRate and VolInc, but GRate was also excluded from the empirical feature space due to the lack of multiple scans per patient, which were required to estimate the growth rate of the tumour). In addition, the tumour grade can often be expressed in terms of the patient's age. It is generally observed that GBM tumours tend to occur more frequently in older patients while low-grade tumours are more common in children [30, 130]. Table 4.1 lists the age, pathology and the tumour growth rate corresponding to each patient in the dataset in order to help more thoroughly explain results given the tumour grade information.

In each experiment, we compare CDM results based on training and classification with each of LGT, NB, and SVM separately.

In Experiment *I* (Table 5.3), we compare the results of UG and GW with CDM results obtained from training each of the three classifiers based on the feature set FS_0, which consists of 75 features and a binary class label (see Table 5.6).

In Experiment *II* (Table 5.4), we compare the results of the three models including CDM results when training classifiers on the feature subset FS_1, which consists only of tumour volume attributes, edema, intensity-based features and a binary class label (see Section 5.2 and Table 5.6).

In Experiment *III* (Table 5.5), we compare the three models' results including CDM's based on the feature subset FS_2, which combines the features in FS_1 with voxel coordinates and template-specific features (see Section 5.2 and Table 5.6).

Given CDM results with three different feature sets, and the results obtained with the standard models, we compare the three models' performance and provide a thorough analysis of CDM's performance. We first discuss the average and best scenarios followed by special cases of tumour diffusion. Section 5.3.4 evaluates the fairness of the dataset and discusses data-related issues that have an impact on CDM's performance. We also compare CDM's accuracy with respect to low-grade versus high-grade tumours. Section 5.3.5 evaluates the training features, and their effect on the model's results. In Section 5.3.6, we discuss issues related to the performance measures.

5.3.2 Average and Best Case Results

In general, it is medically known that edema regions, adjacent to the visibly-enhancing tumour, may contain diffuse cancer cells [61, 72]. Such tumour regions do not enhance at first on T1-weighted images with contrast but as cancer cells build up into small detectable masses, enhancing tumours will appear in regions of edema. Hence, the presence of edema is an indicator of abnormal cell activity possibly due to potential malignant invasion and to the presence of the tumour causing swelling in adjacent normal tissue.

In our data, edema features are associated with the most significant IG scores (see Table 5.2) when measuring information gain on the entire training set for the 17 patients. Consequently, classifiers tend to assign higher probabilities of diffusion to voxels that fall in edema regions, *i.e.*, it is more likely that these voxels will become 'tumour'.

Given the data and the empirical results, cancer diffusion in edema regions is true for at least several patients, *e.g.*, patients #9, 106, and 160; see Figures 5.1, 5.2, and 5.3 respectively for CDM prediction results, using Logistic Regression (LGT) classification based on FS_1 features, as compared to the truth volume. These patients represent average case results where CDM performs more accurately by at least a small percentage as compared

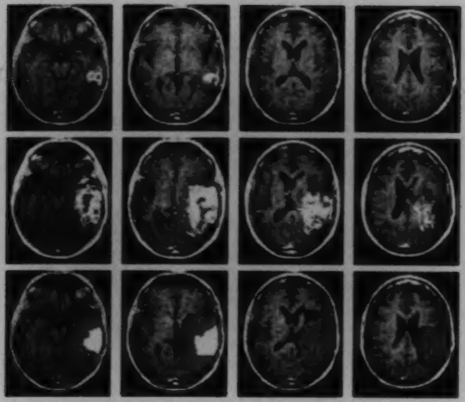

Figure 5.1: Patient #9 image volumes and CDM results

Top: contrast-enhancing T1-weighted image, at the first of two consecutive time points. Middle: the patient's image volume one year later (*i.e.*, this is the truth volume). Bottom: CDM results after the tumour has become nine times larger corresponding to the truth volume (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

to both of UG and GW (see Tables 5.3, 5.4, and 5.5). In these cases, the tumour tends to grow along the edema as glioma cells have already infiltrated into the peritumoural edema regions. These diffuse occult cells did not enhance at first on the T1-weighted images with contrast as these cells may be found in low concentration. But on the next scan of the patient, enhancing tumours appeared in these regions as tumour cells built up into detectable masses.

Infiltration of glioma cells in edema regions is particularly more obvious on the MRI scans for patient #165, which represents the best case results as CDM models tumour diffusion more accurately than UG and GW, by 20% and 12% respectively (see Tables 5.3, 5.4, and 5.5). Figure 5.4 shows CDM prediction results, using LGT classification based on FS_2 feature set, for patient #165 in comparison with the tumour and edema 3.5 months later, after the tumour has tripled in volume.

In average and best case scenarios, the prediction is based on what the classifier recognizes as 'tumour', which are often voxels located in edema regions. Glioma cell infiltration in peritumoural edema may be even more detectable if the truth volume was obtained from a patient scan before that patient would undergo a surgical procedure or would be administered radiation treatment.

5.3.3 Special Cases of Tumour Growth

By 'special cases', we refer to tumour growth scenarios that do not follow usual diffusion patterns along the edema regions or consistent growth behaviour, which is often due to the effect of treatment followed by tumour recurrence. We also refer to scenarios of multifocal tumours (*e.g.*, patient #147, Figure 5.12), non-contiguous enhancing tumours (patient #83,

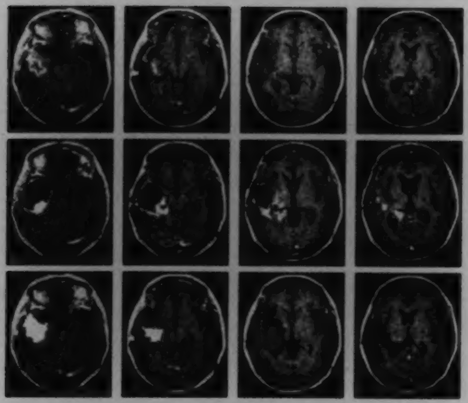

Figure 5.2: Patient #106 image volumes and CDM results

Top: T1-weighted image showing enhancing tumour and peritumoural dark edema regions. Middle: the image volume (the truth) ten months later including a treatment cavity that appears as a dark region in the centre of the tumour. Bottom: CDM results corresponding to the truth volume (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

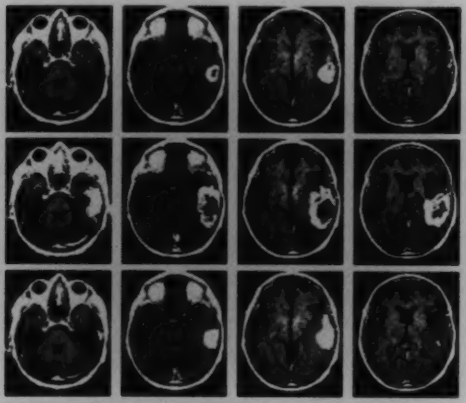

Figure 5.3: Patient #160 image volumes and CDM results

Top: T1-weighted image showing enhancing tumour and dark edema regions. Middle: the image volume (the truth) nine months later. Bottom: CDM results corresponding to the truth volume (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

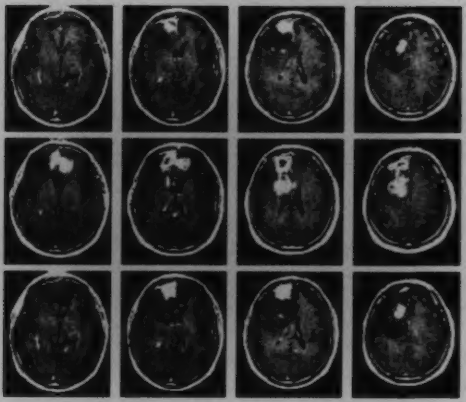

Figure 5.4: Patient #165 image volumes and CDM results

Top: T1-weighted image slices, at the first of two consecutive time points, where the tumour is enhancing and the edema appears as dark regions around the tumour. Middle: the patient's image volume three months later (*i.e.*, this is the truth volume). Bottom: CDM results for the same patient after the tumour has become three times larger. Note tumour growth into edema regions (top row versus the truth volume). CDM prediction volume corresponds to the truth volume (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

Figure 5.9), and treated tumours that include a surgical cavity or an irradiated necrotic region. For example, the truth volumes for patients #106 and 169 include a treatment region that appears as a dark region at the centre of the tumour (Figures 5.2 and 5.13 respectively). Special cases also include shrinkage on one side of the tumour (due to radiation treatment) and fast growth on another side of the same tumour depending on the radiation dose delivered to different regions of the tumour, to peritumoural edema regions, and to the adjacent normal tissue. For example, patient #25 is one scenario of a recurrent tumour that appeared in locations adjacent to the original tumour mass (see Figure 5.6).

Currently, our CDM model does not implement any special handling of these scenarios though we train the model on subjects that have undergone surgical operations and on tumours that appear as multifocal or as non-contiguous enhancing masses on one or both sides of the brain. Our dataset consists of patients that are at different stages of treatment. Not all patients respond to treatment, in particular most GBM tumours continue to aggressively invade adjacent brain tissue even after surgical resection and radiation therapy. It is worth noting that recurrent tumours tend to appear in a new location near the resection cavity or the irradiated tumour volume (*e.g.*, Figure 5.6) . These scenarios are more difficult to predict with our CDM system since our model does not have a method of accounting for tumour recurrence after treatment. For this reason, we have excluded from the training data and from the experiments image volumes where the recurrent tumour appears in different locations across the consecutive scans of the patient (*e.g.*, patient #25, see Figure 5.6).

We tested our model, however, on cases where tumours do not grow along the edema regions or do not show consistent diffusion behaviour along a particular direction. In these cases, CDM's performance was the same or slightly more accurate than UG or GW. Patients #25, 83, and 147 are examples of special tumour growth cases where CDM performed the same as standard models (see Tables 5.3, 5.4, and 5.5).

Patient #6 is one of the scenarios where CDM models the tumour growth more accurately than standard models. In this case, the tumour grows beyond the edema regions visible on the first of two consecutive scans of the patient's brain (see Figure 5.5).

In patient #93, the tumour originated in the left hemisphere of the brain but crossed over to the right hemisphere through the corpus callosum in the posterior section adjacent to the ventricles (see Figure 5.10). In this scenario, CDM modeled the tumour growth more properly than UG and GW, but did not predict diffusion into the right cerebral hemisphere.

In patient #108, the enhancing tumour appeared in some of the edema regions while the rest of the tumour actually continued to grow toward the posterior section of the brain where there was no visible edema on the first of the two consecutive scans of the patient (see Figure 5.11). This tumour diffusion behaviour may be attributed to several factors including the effect of treatment, and the ability of glioma cell masses to enhance in specific brain regions more than others depending on blood supply, and angiogenesis. It is medically known that edema regions may contain non-enhancing glioma cells, but in our experiments, we define the tumour volume as the enhancing tumour regions excluding the edema. In the scenario of patient #108, CDM's accuracy was slightly higher than that for UG and GW.

In summary, given the small training set of 17 patients, and the effect of treatment and glioma recurrence present in most tumour growth scenarios, at least one third of the scenarios can be categorized as 'special cases' of glioma diffusion. While our CDM model does not specifically handle inconsistent diffusion patterns, recurrence, multi-focal tumours, and treatment effect, our model's accuracy is comparable or better than the performance of UG and GW with respect to these scenarios.

5.3.4 Fairness of Dataset

Though we have collected a large number of patient studies for our training dataset, only a small subset of these studies could be used due to the effect of surgical resection and radiation treatment as observed on several patient scans (*e.g.*, Figures 5.6 and 5.12). Also, while we know that all patients are at different stages of treatment, we do not have specific information about the treatment volumes, the radiation dose delivered to different regions of the brain, or patients' response to treatment. The information we have is mostly represented through what we could visually detect – at the human eye level – on the MRI scans.

Because the effect of treatment is present in all patient studies, it is difficult to properly measure the performance of our system given the training data and the tumour recurrence scenarios where the tumour is eradicated from its initial location but appears later in new locations adjacent to the resection region. This effect makes it more difficult for our classification-based model to track tumour growth behaviours across patient scans. It is also difficult to properly evaluate model results since the truth volume (*i.e.*, the tumour at a later time used in evaluating models' performance) may not represent real tumour growth due to the effect of treatment.

Ideally, it would be more desirable if we could train and test our model on patients that have not undergone any treatment. But such data is extremely rare and often unavailable.

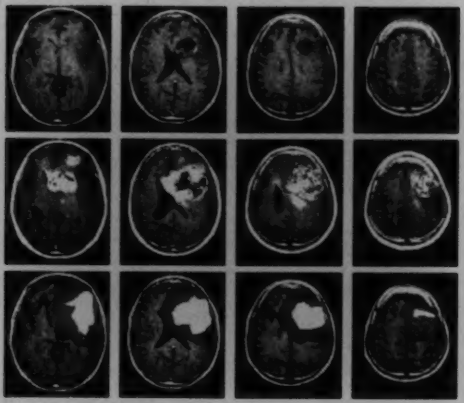

Figure 5.5: Patient #6 image volumes and CDM results

Top: the image volume at the first of two consecutive time points. Middle: the patient's scan six months later (this is the truth volume). Bottom: CDM prediction results corresponding to the truth volume (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

We have compensated for the lack of untreated patient data by excluding from our training set recurrent tumours that appear in different locations across the scans of the same patient. By excluding recurrent tumour instances, we obtain a dataset of consistent tumour growth across the scans of each patient, rather than tumours that decrease in volume then appear in regions adjacent to the original tumour mass. We were also able to use several scans for patients diagnosed with GBM tumours and that did not respond to treatment (*e.g.*, patient #9, Figure 5.1).

Model Performance versus Tumour Grade

Our dataset consists of four different glioma grades ranging from low-grade astrocytomas to the most invasive glioblastomas (GBM). Since GBM is the most common of all primary brain tumours in glioma patients [2, 96, 113], it represents $\frac{2}{3}$ of patients' studies in our data (see Table 4.1). While we have excluded the 'pathology' (*i.e.*, the tumour grade) from the learning features, tumour grade information can be expressed in terms of the tumour's growth rate and the patient's age. There is a direct correlation between the tumour grade and its growth rate. It has also been observed that higher-grade tumours tend to occur more frequently in older patients while low-grade tumours are more common in children [30, 130]. We excluded the growth rate feature from the current feature space because the dataset does not contain multiple studies for many of the patients; these multiple studies were required to estimate the growth rate of the tumour. We were able to include, however, the patient's age in the feature space as it implicitly indicates the tumour grade [30] (see Section 4.5).

Because CDM is a general learning model, it is not restricted to predicting a particular tumour grade, but it requires involving a fair representation of the different tumour types

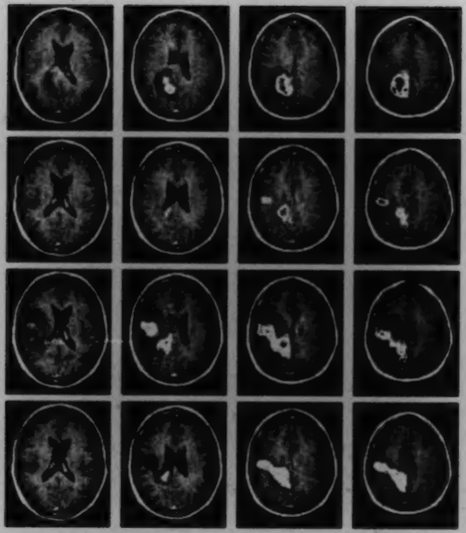

Figure 5.6: Patient #25 image volumes and CDM prediction results
Top to bottom: the image volumes of the same patient at three different time points in chronogical order, followed by CDM prediction results. The images are moving from bottom to top of the patient's head, left to right, on each row. Note treatment effect as both tumour and edema disappeared from the two left slices (top versus second row) and appeared in a different location as seen on the third row. We excluded the top image volume from the training data and the experiments. CDM prediction (bottom row) is based on information of the image volume at the second time point (second row) and on the truth volume (third row). The initial tumour is colored white, true positives are green, false positives red, and false negatives blue. Images at the bottom row (*i.e.*, the prediction) corresponds to the truth images at the third row.

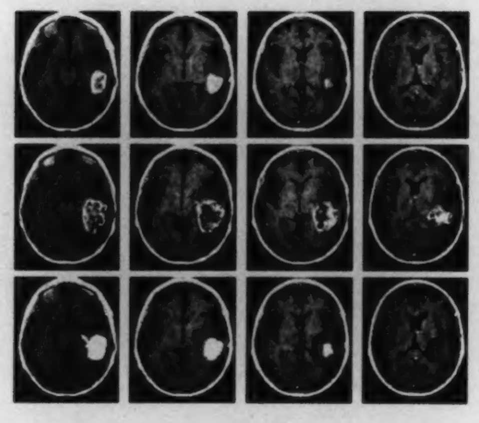

Figure 5.7: Patient #37 image volumes and CDM results

Top to bottom: the image volumes at two consecutive time points followed by CDM prediction results. The initial tumour is colored white, true positives are green, false positives red, and false negatives blue. The images are moving from bottom to top of the patient's head, left to right on each row.

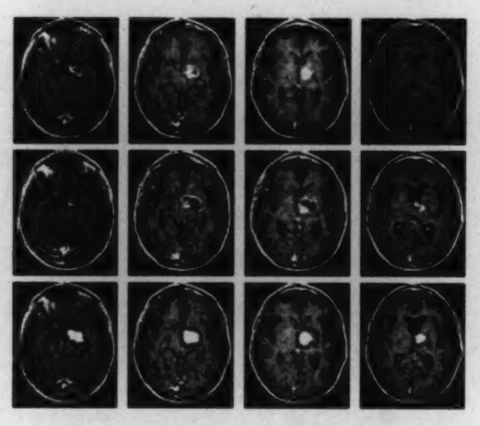

Figure 5.8: Patient #44 image volumes and CDM results

Top to bottom: the image volumes at two consecutive time points followed by CDM prediction results. The initial tumour is colored white, true positives are green, false positives red, and false negatives blue. The images are moving from bottom to top of the patient's head, left to right on each row.

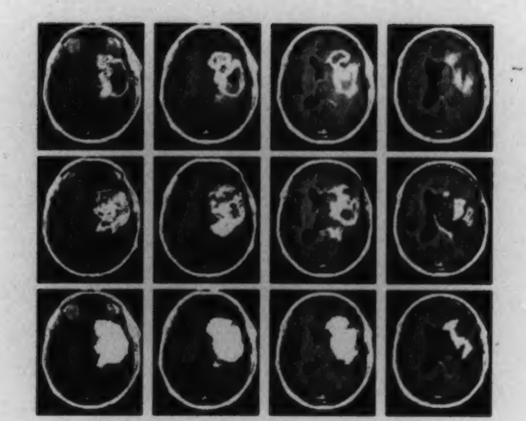

Figure 5.9: Patient #83 image volumes and CDM prediction
Top: the image volume at the first of two consecutive time points. Middle: the patient's scan five months later (the truth volume). Note the tumour did not grow into the anterior brain region along edema regions (left two slices, top versus middle row). The tumour started growing instead toward the posterior region of the brain, detected as non-contiguous masses on the right slice, middle row. Bottom: CDM prediction results evaluated against the truth volume (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

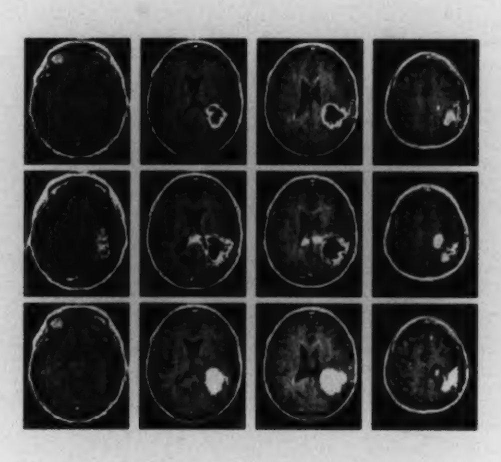

Figure 5.10: Patient #93 scans and CDM results
Top to bottom: the image volumes at two consecutive time points at a four-month interval, followed by CDM prediction results.

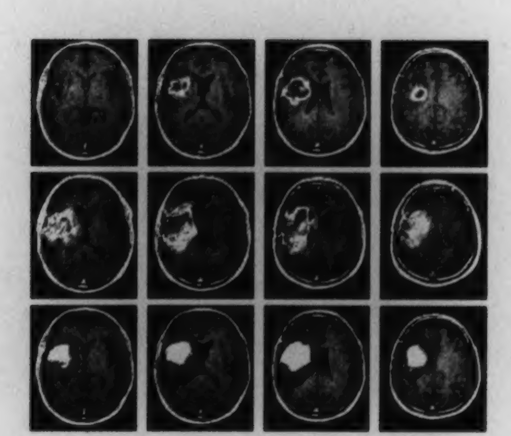

Figure 5.11: Patient #108 scans and CDM results
Top to bottom: the consecutive patient scans at a 12-month interval followed by CDM prediction results evaluated against the truth volume obtained from the second scan (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

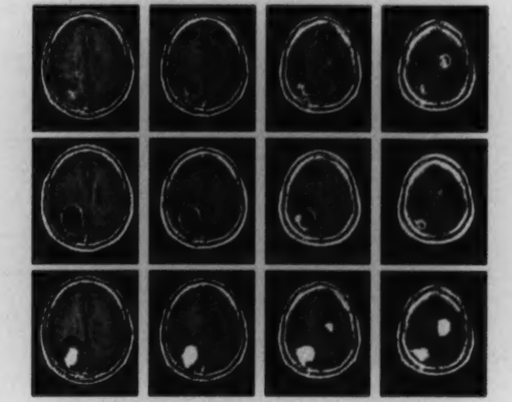

Figure 5.12: Patient #147 scans and CDM results
Top: the image volume at the first of two consecutive time points. Middle: the patient's scan seven months later (used as the truth volume). Note the tumour at each lobe of the brain and the shrinkage of the left lobe tumour due to treatment (top versus middle row). Bottom: CDM's prediction evaluated against the truth volume (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

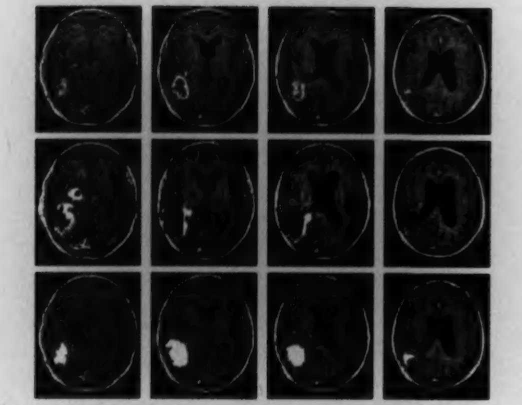

Figure 5.13: Patient #169 scans and CDM results

Top: contrast-enhancing T1-weighted image volume. Middle: The image volume (the truth) a year later showing a necrotic treatment region. Bottom: CDM prediction results corresponding to the truth volume (initial tumour colored white, true positives green, false positives red, and false negatives blue). The images are moving from bottom to top of the patient's head, left to right on each row.

in the learning data in order to properly model all types of gliomas. Currently, low-grade tumours are under-represented in the training data since they are less common than GBM among glioma patients.

Given the representation of tumour grades in the training data, CDM tends to predict GBM tumour growth more accurately than UG and GW, while it performs the same or slightly better than these standard models in the case of low-grade gliomas. We discuss in the following sub-section the statistical confidence in these results. Also, CDM's prediction is based on the probabilities assigned by classifiers to the unlabeled voxels. High-probability voxels are likely to be located in edema regions, in particular in high-grade, large tumours where there is larger edema regions surrounding the tumour (*e.g.*, Figure 5.4). In low-grade gliomas, the edema regions and the tumour volume increase are much smaller than in high-grade tumours (*e.g.*, Figure 5.8) leaving a smaller number of unlabeled voxels for the diffusion model to predict, therefore, a higher error margin in particular at the tumour periphery. This larger error margin is due to a smaller intersecting region of true positives between the prediction volume and the truth volume. Also, the manual segmentation error may be prominent when the tumour volume increase across the patient's scans is small.

Statistical Evaluation of the Three Models using the t-Test

Given the average recall values for CDM (obtained with LGT, SVM, and NB using the feature set FS_1) are 0.598, 0.596, and 0.518, respectively, and the average recall values obtained with UG and GW are 0.524 and 0.566, respectively, we determine if these average recall values are statistically significant by performing Student's t-test [107]. This test indicates whether the difference in the calculated means (*i.e.*, two sets of recall values obtained with two different diffusion models and averaged over 17 patients) may be attributed to "chance".

To report the various possible outcomes: from means not significantly different to means in fact "significantly" different, the probability P that the difference is due to chance is reported. We reject the null hypothesis if P is "small".

We perform the test on paired data, *i.e.*, the same sample of patient studies, to which we applied three different diffusion models (CDM, UG and GW). The paired t-test (Equation 5.1) focuses on the difference between the paired data and reports the probability that the actual mean difference is consistent with no difference. This comparison is aided by the variance associated with the differences.

Let X_i and Y_i be two paired sets of data, each of sample size n, X_{avg} and Y_{avg} are the mean values for the two sets respectively, and X_{s_i} and Y_{s_i} are the standardized sample means such that $X_{s_i} = X_i - X_{avg}$ and $Y_{s_i} = Y_i - Y_{avg}$. The t value is computed as the following:

$$t = (X_{avg} - Y_{avg})\sqrt{\frac{n(n-1)}{\Sigma_{i=1}^{n}(X_{s_i} - Y_{s_i})^2}} \tag{5.1}$$

We perform the t-test by first computing the t values for each of the result samples obtained with the three diffusion models, CDM, UG and GW. We use $n-1$ degrees of freedom and 95% confidence interval for the mean. We then compute the probability assuming the null hypothesis (*i.e.*, the means are not significantly different). The standard deviation measured for each of UG and GW is 0.126 and 0.117 respectively. The standard deviation for CDM results obtained with each of LGT, SVM, and NB is 0.129, 0.128, and 0.117 respectively. We perform the test on the following pairs of data:

- CDM results obtained with LGT using feature set FS_1 versus UG results: the t value is 4.14, and the probability, assuming the null hypothesis, is 0.001. In this case, we reject the null hypothesis and we conclude that the average recall values obtained with CDM and UG are significantly different.

- CDM results obtained with LGT using feature set FS_1 versus GW results: the t value is 3.61, and the probability, assuming the null hypothesis, is 0.002. Therefore, we reject the null hypothesis and we conclude that the average recall results obtained with CDM and GW are significantly different.

- CDM results obtained with LGT versus SVM, using FS_1: the t value is 1.09, and the probability, assuming the null hypothesis, is 0.293. In this case, we do not reject the null hypothesis, which means that results obtained with LGT and with SVM are not significantly different.

- CDM results obtained with LGT versus NB, using FS_1: the t value is 2.84, and the probability, assuming the null hypothesis, is 0.012, which leads to rejecting the null hypothesis. Results obtained with a generative classifier (*i.e.*, NB) can be significantly different from those obtained with discriminative classifiers (*i.e.*, LGT).

5.3.5 Evaluation of Training Features

Currently, most of our training features are based on image intensities and voxel coordinates as well as tumour volume information. These attributes are obtained from *normal* versus *abnormal* regions visually detected on the MRI scans and are therefore susceptible to subjective human interpretations of the definition of *abnormality*. The current features have offered an important but limited contribution given training data limitations and treatment effect. In more consistent tumour growth scenarios (Section 5.3.2), the training features incorporated in our CDM model helped more accurately predict tumour growth than the standard models UG and GW. But in other cases (Section 5.3.3), CDM performed the same as the standard models.

In addition, when we train and test on the same data, as opposed to testing with cross-validation, we do not obtain perfect accuracy. Applying classifiers on the feature set FS_0 to train and test on the same data yielded a precision of 0.711 and a recall of 0.849 with Logistic Regression, and 0.707 precision, 0.714 recall with Naïve Bayes. These non-perfect precision and recall values are due to overlapping data points between the 'tumour' and 'non-tumour' classes. In other words, the attributes of a voxel may indicate that this voxel will be labeled as 'tumour' but on the truth volume, the voxel may fall in the 'non-tumour' region. This effect might be due to treatment and to other factors given our training data is limited to information visually detected on the MRI scans, and given the error margin at the boundary of the segmented tumour volumes. In addition, our current training features may not be sufficiently expressive of the *abnormal* regions – *i.e.*, regions of potential glioma diffusion. It is also worth noting that information acquired from the MRI scans does not help differentiate between increased water content and micro-infiltration of glioma cells in peritumoural edema regions. That is, these edema regions likely contain diffuse glioma cells that have infiltrated through normal tissue but can be found, with MR spectroscopy, in low concentrations below the threshold of detection of MR imaging.

We believe that model accuracy might have significantly improved if we could include in our features information acquired through spectroscopy, metabolic tumour data, and descriptions of the genetics of the tumour cells. As opposed to visually detecting abnormal regions on the MRI scans, spectroscopy helps indicate invasion of cancer cells into adjacent normal tissue even though these cells may be found in very low concentrations and are undetected on the MR images. Metabolic and genetic tumour features may help predict malignant diffusion in tumour regions where the cancer cells are more active and proliferate at a faster rate than in the rest of the tumour. On the MRI scans, this effect might be detected as fast tumour growth on a particular side of the tumour border while the cells at the rest of the tumour periphery proliferate with a lower rate.

At the present time, it may be very costly to obtain spectroscopy and genetic tumour information. Most existing approaches use therefore CT or MRI scans to obtain model parameters and to evaluate results.

Comparison of Results using Different Feature Sets

Currently, training with different subsets of feature combinations (*i.e.*, FS_0, FS_1, and FS_2 described in Section 5.2) does not yield significantly different results (see Tables 5.3, 5.4, and 5.5). In some scenarios, the accuracy may increase or decrease by $2-3\%$ depending on the features used in the training and prediction, and on whether these scenarios require a higher level of generalization.

For example, the accuracy slightly increased, for patients #6, 37, 44, and 83, when training a Logistic Regression classifier (LGT) with FS_1 – which excludes voxel coordinates, tissue map, template intensities, and brain tissue types – versus training with the FS_0 feature set (compare Tables 5.3 and 5.4). This slight increase in accuracy may be attributed to the higher level of generalization provided in training with FS_1 features. Also, tumour-induced pressure on the ventricles is common in these four scenarios, suggesting that it is more beneficial not to learn template-specific information in particular in the ventricles region (*i.e.*, the template's ventricles may overlap with tumour growth regions on the patient's image), and brain tisue information where the tumour has pushed the CSF. In cases where the tumour volume is significantly large or pushes CSF anatomic regions, SPM does not properly segment the grey and white matter and CSF due to the size and location of the tumour. In other scenarios, *e.g.*, patients #8, 108, and 147, the accuracy slightly decreased, likely due to the treatment effect (tumour shrinkage and recurrence) that becomes more prominent when training mainly with edema features (as in FS_1). In these scenarios, the classifiers assign higher probabilities to voxels situated in edema regions, assuming potential tumour invasion along the edema, but since the enhancing tumours may appear in only some of these edema regions (due to the effect of treatment and to other anatomical factors), this prediction will result in a large number of false positives.

Training with FS_2 feature set yielded similar results to training with the entire feature space (compare Tables 5.3 and 5.5). This is because FS_2 incorporates the template-specific information and voxel coordinates that were excluded from FS_1.

Comparison of Classifiers' Performance

By comparing the results of the three classifiers (Tables 5.3, 5.4, and 5.5), we observe that the results of Logistic Regression (LGT) are more or less comparable to those of SVM while both classifiers are performing more accurately than Naïve Bayes (NB).

In our results, the average recall for NB results tends to be lower than the average recall obtained with LGT and SVM, which means than NB classification yields a larger number of false negatives (*i.e.*, it labels more 'tumour' voxels as 'non-tumour'). In each of the three experiments, we average the recall values obtained for the 17 test patients. The average recall for the results in Table 5.3 is 0.594, 0.565, 0.592 for LGT, NB, and SVM respectively. In Table 5.4, the average recall is 0.598, 0.518, 0.596 and in Table 5.5, it is 0.594, 0.544, 0.592, for LGT, NB, and SVM respectively. Note the average recall for NB is $3-8\%$ lower in the three experiments.

We attribute NB results, in comparison with LGT results, to the generally observed performance of generative (*i.e.*, NB in our experiments) versus discriminative (*i.e.*, LGT) classifiers. Given a target function $P(Y|X)$ that we wish to learn, where Y is a target attribute and X is the instance space, a *generative* classifier views the distribution $P(X|Y)$ as describing how to generate random instances X conditioned on the target attribute Y [82]. A *discriminative* classifier directly estimates $P(Y|X)$ as it views the distribution $P(Y|X)$ as directly discriminating the value of the target attribute Y for any given instance

X [82]. Naïve Bayes classifiers have a higher asymptotic error (as the number of training examples becomes large) than Logistic Regression models [86]. In the general case, NB and LGT converge to their asymptotic accuracies at different rates in the order of *log* n and n respectively where n is the dimension of X, *i.e.*, LGT converges more slowly [86, 82]. With a large number of training examples, the error of the LGT classifier is smaller than that of the generative NB, which yields better performance with LGT [86, 82]. This effect has been observed in our results as LGT yielded a higher average recall for the 17 test patients as compared to NB results.

5.3.6 Fairness of Performance Evaluation

While current approaches compare their model results by measuring the distance in millimeters between the boundaries of the predicted tumour and of the truth (see *e.g.*, [21] and [138]), we use precision and recall measures to evaluate CDM performance at the voxel level. These measures are more accurate in assessing system performance as opposed to graphically measuring the distance between the prediction results and the truth (see Figure 2.1).

We note, however, that the results of the three models, CDM, UG and GW, include a non-quantified error margin at the boundaries of the manually segmented tumour volumes, due to human error and to the definition of tumour versus normal regions.

Presently, CDM's performance is restricted by our definition of *tumour volume*, which consists of the enhancing tumour and abnormal textures adjacent to the enhancing volume. This definition does not generally include the peritumoural edema regions. In scenarios where CDM predicts glioma diffusion into edema regions, the prediction accuracy will depend on whether enhancing tumours will appear in the edema regions on the next scan of the patient. In these cases, the accuracy might be under-estimated since edema regions are likely to contain diffuse glioma cells. But given our definiton of the abnormal regions, we evaluate the prediction in terms of enhancing tumours only.

CDM's performance is also limited by the tumour volume increase. In smaller tumour volumes (often low-grade gliomas), the error margin at the tumour boundary may represent a larger fraction of the total number of unlabeled voxels. CDM tends to model larger tumours more accurately.

Another limitation is the spatial interpolation step (see Chapter 3), which also includes a non-measured error margin at the tumour boundary. After spatial interpolation of the tumour volumes, we discard low-intensity voxels that fall below the 50% threshold (obtained from dividing the voxel's intensity by the maximum intensity). The output of the registration and spatial interpolation steps is a 91-slice image volume obtained from a 20-slice image volume (before processing). Errors observed in the output volume include tumour voxels overlapping with the petrous bone, and other interpolation artifacts that appear as non-smooth lines or sudden intensity changes across the slices of the output image (as described in Section 3.10).

5.4 Conclusion

In this chapter, we have discussed the challenges involved in evaluating the performance of our classification-based model, CDM. We have also discussed the feature selection, issues related to training data, and the evaluation measures used in assessing system performance. We then presented experiments performed with three different classifiers and the results obtained when training with three sets of feature combinations.

We conclude from the results that, though glioma diffusion prediction is a rather challenging problem, it is feasible to model glioma growth based on learning and classification more accurately than with standard methods. Several scenarios in the experiments showed that our classification-based method can 'track' glioma diffusion patterns, in particular when the enhancing tumour spreads along the edema regions. Peritumoural edema is due to both increased water content and to glioma cell infiltration into normal tissue. The presence of edema may therefore indicate abnormal cell activity as glioma cells continue to build up in the edema into enhancing, detectable tumour masses [61, 72].

We also conclude that glioma diffusion does not conform with radial uniform growth (performed by the UG model) but tumours tend to grow into asymmetric volumes. In addition, several factors are involved in this diffusion process, and it is not sufficient to account only for the heterogenous brain tissue when modeling tumour growth (as performed with the GW growth model). Instead, to successfully model glioma growth, it is important to incorporate information specific to the voxels adjacent to the tumour, the edema regions, the brain anatomy and to the patient. It is medically observed that edema regions likely contain diffuse glioma cells [61, 72], and that tumours may continue to grow along the direction of the white fibres [113, 90]. But these tumour cells do not diffuse into CSF regions (*e.g.*, the ventricles) though the tumour-induced pressure may deform these anatomic structures. An accurate diffusion model should be able to differentiate between the heterogeneous brain anatomy. Our CDM model utilizes template-based information to allow tumour growth into the grey and white matter only, and to simulate tumour-induced pressure on CSF regions at a 10% diffusion rate.

To evaluate the statistical confidence in our conclusions regarding the performance of the three models, we performed Student's t-test [107] (see Section 5.3.4). The results of this statistical test showed that the average recall values, obtained with the three different models (CDM, UG and GW), are significantly different in that these differences are unlikely to be due to chance. We applied the test to the recall values (averaged over the 17 patients) given that our implementation of the three models yields equal precision and recall values (see Section 4.6).

Our learning system, CDM, is also efficient as in most cases, it performs the diffusion modeling in real time. Given a preprocessed MRI scan and a segmented tumour, CDM performs tumour growth modeling, using an already learned classifier, in terms of minutes depending on the additional number of tumour voxels to grow (*i.e.*, the target volume of the tumour). The preprocessing of an MRI scan of a single patient and the feature extraction are performed in approximately one hour. The average efficiency of the diffusion algorithm, given the labels predicted by Logistic Regression based on the feature set FS_0, is 9.7 minutes averaged over the 17 patients.

We believe that CDM would have performed more accurately if the effect of treatment was not present in the training data, in particular in special case scenarios where tumour shrinkage and recurrence were observed. In these cases, CDM's performance is comparable to that of standard models.

Also, CDM tends to perform more accurately when predicting diffusion for high-grade gliomas than low-grade and MOA tumours (Section 5.3.4). This is because these lower-grade gliomas are not properly represented in the training data (about $\frac{1}{3}$ of our training instances), and they are more likely to respond to treatment in comparison with high-grade gliomas. CDM also performs more accurately in predicting large tumours, where signs of abnormality (edema and texture) are detectable more easily on the MRI scans. Since the model's prediction is based on information acquired from the brain images, including image intensities and edema regions, these features have an important role in the model's accuracy.

Given the three different experiments presented in this chapter, empirical results showed that combining different features did not have a significant effect on CDM's performance in general. These feature combinations have helped slightly increase the accuracy in glioma diffusion scenarios that require a higher level of generalization, but did not increase the overall average accuracy. It should also be mentioned that predictions based on NB classification yielded lower recall than with LGT and SVM. This means that NB classification yields a high number of false negatives resulting in a larger number of iterations with the diffusion algorithm, causing the algorithm to terminate before reaching the target volume. NB classifiers have been more sensitive to ambiguous training data where some 'tumour' versus 'non-tumour' voxels may have similar attributes in particular if located in the edema regions. LGT and SVM yielded similar results though LGT performed far more efficiently, at training time, as compared to SVM that required further sub-sampling of the training data to reduce training time.

In summary, CDM performs more accurately in most scenarios as compared to radial uniform growth (UG). It is also more accurate than diffusion based on grey and white matter (GW) in several scenarios. Currently, CDM's prediction is mainly based on features extracted from the MRI scans and can only be as accurate as the threshold of detection of tumour regions with MR imaging and the segmentation of tumour volumes.

Through this research study, we hope to help radiation oncologists identify regions where there may be potential glioma invasion in order to specify more precisely the treatment volume and radiation dosage delivered to the patient's brain; hence, decreasing the chance of malignant recurrence.

Feature Name	IG score	F. Name (Cont'd)	IG Score
DistArea	0.03798	Age	0.01292
EdemaNei3	0.03565	GmNei5	0.0107
EdemaNei6	0.03427	GmNei1	0.01068
EdemaNei5	0.0341	GM	0.01049
EdemaNei1	0.03326	GmNei4	0.00989
EdemaNei2	0.03274	GmNei2	0.00984
EdemaNei4	0.02968	GmNei6	0.00984
Edema	0.02923	GmNei3	0.00891
T2IntNei3	0.02733	Ch2WmNei3	0.00805
T1Int	0.02699	Ch2WmNei6	0.00778
TissueMap	0.0238	Ch2WmNei5	0.00755
Ch2Int	0.02168	Ch2WM	0.00745
T2IntNei6	0.02136	y coord	0.00716
EuclDist	0.02113	Ch2WmNei1	0.00707
T2Int	0.02111	Ch2WmNei2	0.00701
T2IntNei5	0.02095	Ch2WmNei4	0.00643
T2IntNei1	0.02013	Ch2CsfNei3	0.00612
VolInc	0.01994	Ch2CsfNei5	0.00574
T2IntNei2	0.01989	Ch2CSF	0.00527
x coord	0.01969	Ch2CsfNei6	0.00515
Ch2GmNei3	0.01932	Ch2CsfNei1	0.00497
Ch2GmNei6	0.01887	Ch2CsfNei2	0.0046
T1cInt	0.01885	WmNei6	0.00439
Ch2GM	0.01883	WmNei5	0.00436
Ch2GmNei5	0.01845	Ch2CsfNei4	0.00432
Ch2GmNei1	0.01836	WM	0.00428
Ch2GmNei2	0.01814	WmNei2	0.00424
T1cIntNei1	0.01761	WmNei1	0.00421
T1cIntNei5	0.01753	WmNei3	0.0041
T2IntNei4	0.01727	WmNei4	0.00397
T1cIntNei6	0.01705	CsfNei2	0.00307
T1cIntNei3	0.01694	CsfNei4	0.00294
EdemaPerc	0.01669	CsfNei1	0.00287
T1cIntNei2	0.01646	CSF	0.00276
Ch2GmNei4	0.01598	CsfNei5	0.00259
z coord	0.01488	CsfNei6	0.00236
AreaVol	0.014	CsfNei3	0.00146
T1cIntNei4	0.01332		

Table 5.2: Information Gain (IG) scores ranked in descending order for the FS_0 feature set, a subset of the feature space, described in Section 4.5

Patient Id	CDM(LGT)	CDM(NB)	CDM(SVM)	UG	GW
6	0.703	0.617	0.704	0.660	0.635
8	0.633	0.544	0.620	0.611	0.588
9	0.765	0.807/0.601	0.765	0.679	0.730
25	0.466	0.463	0.468	0.390	0.452
37	0.628/0.602	0.647/0.620	0.624/0.598	0.476/0.456	0.661/0.634
44	0.544	0.536	0.543	0.564	0.579
51	0.575	0.552	0.574	0.448	0.553
69	0.507	0.515	0.510	0.509	0.523
83	0.365/0.371	0.364/0.371	0.368/0.375	0.362/0.369	0.350/0.354
93	0.726	0.718	0.721	0.674	0.676
105	0.433	0.419	0.434	0.431	0.377/0.373
106	0.723	0.683	0.721	0.476	0.685
108	0.676	0.665	0.673	0.630	0.654
147	0.361	0.378	0.373	0.258	0.367
160	0.692/0.712	0.637/0.655	0.681/0.701	0.649/0.668	0.628/0.646
165	0.649	0.663	0.638	0.469	0.544
169	0.644	0.635	0.646	0.624	0.629

Table 5.3: Experiment *I* Results

Precision and Recall for the three models with three classifiers for CDM using the feature set FS_0 (described in Section 5.2 and in Table 5.6). For efficiency purposes, data used in training SVM has been sub-sampled by including every 15^{th} voxel.

Patient Id	CDM(LGT)	CDM(NB)	CDM(SVM)	UG	GW
6	0.720	0.788/0.469	0.723	0.660	0.635
8	0.614	0.422/0.326	0.608	0.611	0.588
9	0.765	0.952/0.383	0.765	0.679	0.730
25	0.465	0.463	0.468	0.390	0.452
37	0.659/0.631	0.657/0.630	0.653/0.626	0.476/0.456	0.661/0.634
44	0.592	0.486	0.588	0.564	0.579
51	0.571	0.559	0.568	0.448	0.553
69	0.504	0.484	0.507	0.509	0.523
83	0.378/0.385	0.373/0.380	0.374/0.381	0.362/0.369	0.350/0.354
93	0.716	0.706	0.715	0.674	0.676
105	0.435	0.431	0.427	0.431	0.377/0.373
106	0.730	0.736	0.725	0.476	0.685
108	0.663	0.581/0.554	0.667	0.630	0.654
147	0.343	0.279	0.353	0.258	0.367
160	0.704/0.724	0.715/0.673	0.702/0.722	0.649/0.668	0.628/0.646
165	0.656	0.652	0.644	0.469	0.544
169	0.648	0.595	0.651	0.624	0.629

Table 5.4: Experiment *II* Results

Precision and Recall for the three models with different classifiers for CDM using FS_1 features (described in Section 5.2 and in Table 5.6). For efficiency purposes, data used in training SVM has been sub-sampled by including every 15^{th} voxel.

Patient Id	CDM(LGT)	CDM(NB)	CDM(SVM)	UG	GW
6	0.704/0.705	0.752/0.489	0.705	0.660	0.635
8	0.632	0.465	0.614	0.611	0.588
9	0.765	0.940/0.506	0.765	0.679	0.730
25	0.467	0.466	0.468	0.390	0.452
37	0.638/0.612	0.648/0.621	0.625/0.599	0.476/0.456	0.661/0.634
44	0.542	0.503	0.546	0.564	0.579
51	0.571	0.561	0.575	0.448	0.553
69	0.506	0.516	0.511	0.509	0.523
83	0.364/0.371	0.369/0.376	0.364/0.371	0.362/0.369	0.350/0.354
93	0.723	0.715	0.723	0.674	0.676
105	0.443	0.418	0.434	0.431	0.377/0.373
106	0.719	0.706	0.716	0.476	0.685
108	0.675	0.643	0.673	0.630	0.654
147	0.354	0.338	0.361	0.258	0.367
160	0.689/0.709	0.655/0.674	0.682/0.704	0.649/0.668	0.628/0.646
165	0.664	0.645	0.657	0.469	0.544
169	0.644	0.606	0.649	0.624	0.629

Table 5.5: Experiment *III* Results

Precision and Recall for the three models with different classifiers for CDM using FS_2 features (described in Section 5.2 and in Table 5.6). For efficiency purposes, data used in training SVM has been sub-sampled by including every 15^{th} voxel.

Feature Set	List of Features
IFS	*Patient:* Age, pathology, history, KPS
	Anatomical: AAL & Brodmann labels
	Tumour: TumVol, EdemaVol, AreaVol, GRate, EdemaPerc, VolInc, VolDiff
	Voxel: WM, GM, CSF, Edema, EuclDist, T1Int, T2Int, T1cInt, DistArea, DistVol
FS_0	*Patient:* Age
	Tumour: AreaVol, EdemaPerc, VolInc
	Voxel: coordinates, TissueMap, EuclDist, DistArea, Edema, WM, GM, CSF, Ch2Wm, Ch2GM, Ch2Csf, T1Int, T2Int, T1cInt, Ch2Int
	Neighbourhood: EdemaNei, T2Int, T1cInt, WmNei, GmNei, CsfNei, Ch2WmNei, Ch2GmNei, Ch2CsfNei
FS_1	*Patient:* Age
	Tumour: AreaVol, EdemaPerc, VolInc
	Voxel: EuclDist, DistArea, Edema, T1Int, T2Int, T1cInt
	Neighbourhood: EdemaNei, T2IntNei, T1cIntNei
FS_2	*Patient:* Age
	Tumour: AreaVol, EdemaPerc, VolInc
	Voxel: coordinates, TissueMap, EuclDist, DistArea, Edema, Ch2Wm, Ch2GM, T1Int, T2Int, T1cInt, Ch2Int
	Neighbourhood: EdemaNei, T2IntNei, T1IntNei, Ch2WmNei, Ch2GmNei

Table 5.6: The four feature subsets used in the experiments

List of features corresponding to each feature subset derived from the potential feature space listed in Section 4.5. Identifiers (feature names) used in this table are the same as in Section 4.5.

Chapter 6

Conclusion

In previous chapters, we have introduced the problem of modeling glioma diffusion based on information acquired from MR images. This problem has been motivated by its practical application in radiation therapy, and its potential usefulness in improving current treatment methods. Given the characteristics of the problem, glioma diffusion modeling can be solved as a pattern recognition task where machine learning algorithms learn glioma diffusion patterns from the attributes of the voxel regions adjacent to the tumour volume. In Chapter 3, we have presented the processing pipeline of the imaging data sequences, and provided an overview of the model's framework. Chapter 4 defined the problem, and presented the implementation of the proposed diffusion model, CDM, along with an implementation of two standard models. Chapter 5 evaluated the performance of the proposed model in terms of patient-specific testing. This final chapter summarizes the contribution of the thesis, and discusses potential future directions of research that directly follow from this work.

6.1 Contribution

While most glioma growth models proposed in the state-of-the-art literature are based on mathematical approaches that predict diffusion at the macroscopic scale, as described in Chapter 2, this thesis presents a general model based on learning and classification at the voxel level. Our proposed model learns glioma diffusion patterns from MR image volumes acquired for 17 patients, and classifies the unlabeled voxels of a new patient based on the attributes of these voxels. Voxel attributes are obtained from image-based features, information specific to the tumour volume, characteristics of the brain tissue in regions adjacent to the tumour, and attributes of the patient. A diffusion algorithm is then initialized from the tumour volume observed on the patient's MRI scan at one time. The algorithm iterates around the tumour border adding candidate voxels – *i.e.*, voxels that are more likely to become tumour as predicted by the classifier – until the tumour reaches a specified volume. To evaluate this approach, we set the target volume to the volume observed at a later scan for the patient, and we then compare the predicted volume with the actual tumour volume observed at this second scan.

Results show that image-based features, in particular intensity-based and edema attributes, help 'track' glioma diffusion in adjacent tissue. The results also show that glioma cells often infiltrate through peritumoural edema regions though these tumour cells may not enhance at first on T1-weighted MR image with contrast. Our proposed model is able to recognize these diffusion patterns and is able to model glioma growth in these cases more accurately than standard methods. In other scenarios, particularly recurrent gliomas that appear in new locations near the treated volume, CDM performs the same as standard models. In these special scenarios, image-based features do not provide sufficient information that would help track the activity of occult glioma cells. In some cases, cancer cells may already have infiltrated through adjacent normal tissue without any detectable signs of abnormality on the MRI scans. In these scenarios, information solely obtained from the MR images will not provide the necessary features required for recognizing *abnormal* tissue regions where glioma cells have infiltrated but are found in very low concentrations below the threshold of detection of MR imaging.

In addition, the comparison of results obtained from the three models suggests that glioma cells do not follow a radial uniform growth pattern, but tend to be more invasive to normal tissue in particular regions of the brain, *i.e.*, regions that represent "highways" for glioma cell diffusion. Glioma diffusion is based on a large number of factors, some of which may still be unknown, but include the tumour grade or histology, the brain anatomical structures, patient attributes, and properties of the tissue adjacent to the diseased region. Unlike radial uniform diffusion conventionally assumed in cancer treatment and growth models based only on the diffusion rates in white versus grey matter [110], our proposed model is more sophisticated. That is, CDM takes into account a combination of factors – mainly available through the MRI scans – that are currently known to have a role in glioma invasion, and uses the information from the MRI scans to produce more accurate results as compared to standard approaches.

6.2 Future Directions

Given the current results and the special case scenarios, there are several future directions that are worth exploring both at the data processing level and in the implementation of the diffusion model. Section 6.2.1 presents potential improvements that can be made at each step of the data processing pipeline and in the model framework overall. Section 6.2.2 suggests additional data required to improve the performance of the diffusion model, and future work based on this data. Section 6.2.3 suggests additional features and classification methods that represent promising future research directions that may help discover glioma diffusion patterns. Finally, we present in Section 6.2.4 potential additions to the model to help differentiate brain anatomical regions and to guide the diffusion algorithm into regions where gliomas are more likely to grow.

6.2.1 Framework Improvements

Future improvements can be made in each step of the data processing pipeline, and include modifications or additions to the noise reduction step, the spatial registration, the intensity standardization, the delineation of the tumour volumes from the brain images, and the segmentation of brain tissue classes.

Noise reduction can be improved by incorporating an additional step to reduce local noise at the 2D and 3D levels, *e.g.*, using noise reduction filters such as *Smallest Univalue Segment Assimilating Nucleus filter* [99] (suggested in [95]), which preserves the data in the image while reducing the effects of local noise. Local noise is the corruption of signal recorded at each pixel; it is independent of pixel location, but dependent of the tissue measured at this location. Other improvements include making the noise reduction algorithms more robust to abnormalities in the image volumes, and to use algorithms that make the transformed intensities more consistent across the scans of different patients. This applies to the intensity standardization as well. Also, a non-linear intensity standardization algorithm or classification-based methods are likely to be more effective than existing methods.

Spatial registration can be made more accurate by accounting for abnormalities in patient images, different scanning angles and the slice positioning along the z axis in comparison with the template image. Another registration step that can be added to the processing of the MRI scans is the coregistration of images from different modalities to ensure that T1-weighted and T2-weighted images of the same patient properly align. Coregistration can be applied before the tissue segmentation step since both image modalities are used in segmenting brain tissue into three separate classes.

Tissue segmentation also needs to consider the abnormal tumour regions in the patient's image since the current output tissue volumes include a corrupted region around the tumour. Also, improving the intensity inhomogeneity reduction, which is applied prior to tissue segmentation, will likely help improve the differentiation between the tissue classes.

In the tumour segmentation step, it would be more desirable to use an accurate automatic method (rather than manual tumour segmentation). An automatic segmentation is more likely to make the delineated tumour volumes more consistent across patients' images, in particular at the tumour boundaries, where manual segmentation generally introduces an error margin. Making tumour segmentation more accurate requires proper discrimination of the abnormal versus normal regions in the image, which would in turn help improve the results of our model since the classification, prediction and model evaluation tasks are subject to the definition of *abnormality* detected on the MRI scans.

In addition to improving each individual step in the processing pipeline, the overall framework can also be modified by adding or removing one or more of the above steps, combining several methods in one task (such as intensity inhomogeneity reduction and tissue segmentation), and changing the order in which the processing steps are applied to the MRI data.

6.2.2 Data-related Suggestions

In order to improve the accuracy of the proposed diffusion model, many data-related issues (*e.g.*, noise, time-interval inconsistency across scans, and tumour volume variations) need to be resolved since supervised learning methods are generally sensitive to the training data. First, we need to obtain a larger training dataset, which would require access to a significantly large data collection, in order to select from this collection usable studies. It is also important that this data collection would contain a varied population of patients from different age categories and with a range of tumour grades. The dataset should also include tumour occurrences in different locations of the brain. Because it may not be possible to acquire data for untreated patients, it would be useful to obtain treatment information including the type of treatment and the radiation dose (which may be used as features), in order to account for the treatment effect in the learning task.

Once a large training dataset is available, training on the same tumour type (or histology) may help improve the results in particular in the case of low-grade gliomas, which are currently under-represented in our training data. It may also be worth experimenting with different definitions of the *abnormal* region. Currently, the prediction task and the model evaluation are mainly based on 'enhancing tumours' as the 'abnormal' region. Other definitions of *abnormality* may also include peritumoural edema regions assuming that glioma cells have infiltrated into these regions. But in order to include the peritumoural edema into the definition of *abnormality*, we must have access to a very large set of patients' studies such that the tumour volume increase between each two consecutive studies of the same patient should be large enough to make a prediction.

Other useful additions to the data include using other image modalities, *e.g.*, FLAIR, or even a different type of scan, *e.g.*, PET (Positron Emission Tomography) that has a lower threshold for finding low-concentration cancer cells as compared to MRI. It would also be useful to incorporate DTI data, which can help track the direction of glioma diffusion. Combining different image modalities or data from different scanning techniques can provide further insight into glioma diffusion behaviour and potential regions where cancer cells are more likely to infiltrate.

6.2.3 Suggestions for the Classification Task

In addition to improvements of the model framework and the training data, we discuss future directions with respect to the machine learning task and the diffusion model. An obvious addition to our current system is a broader feature space that would not only include image-based and volume-specific features, but also attributes of the cancer cells at the microscopic level. These features would include metabolic data of the tumour, the interaction between normal and cancer cells, and the genetic features of the tumour's cells. This is particularly useful when predicting the growth behaviour of tumours that progress from low-grade to high-grade, and of tumours that have different types of cells (*i.e.*, MOA). Another useful addition is to incorporate MR spectroscopy data, which helps indicate more precisely the locations of glioma diffusion into adjacent tissue. Incorporating the patient's genetic history, brain anatomy attributes, and other patient-specific data, may also help improve results. In addition, as mentioned in Section 6.2.2, including features from different MR modalities or other imaging techniques (DTI and PET) can help more accurately define the abnormal regions and identify occult cell infiltration, in particular, at the tumour periphery where it is difficult to identify a clear boundary between the tumour and the normal tissue. Another type of features that has not been yet considered is textural features, which can be derived by averaging the image intensities or by computing intensity differences between the images. This can be achieved using several image modalities, spatial priors (*e.g.*, the templates and brain mask prior), and different types of brain scans. Other features that may be worth considering are the attributes of the voxels that are diagonally adjacent to a training voxel (currently, we only consider attributes of the voxels that are directly adjacent to training voxels). Also, while we potentially consider using a single growth rate feature that describes the diffusion rate of the tumour as a whole, the tumour volume can instead be partitioned into several volumes near the tumour periphery (in terms of cubic voxels), and different growth rates can then be computed for each of these partitions. This is because gliomas do not diffuse at an equal rate in all directions alike, but tumours tend to grow faster in particular regions given that the action of diffusion is always near the periphery of the tumour, while the cells at the centre do not proliferate.

In addition to incorporating new features, investigating different feature selection methods will help identify the best feature combinations; in particular, using an automated feature selection method can result in performance gains, given that current results show that combining feature sets does not necessarily improve the model's accuracy.

The characteristics of the glioma diffusion problem suggest that using classification based on neighbourhood interdependency would also be a promising future direction. 'Interdependency' refers to the progressive diffusion of malignant cells into a particular voxel given the already existing glioma cell masses adjacent to this voxel. To account for voxel interdependencies, the diffusion algorithm should be implemented in a way that the prediction of the voxels at iteration i depends on the prediction of voxels at iteration $i-1$ starting from the tumour border. In this case, the classifier will not label all the voxels of a new patient at once prior to running the diffusion algorithm, but it will label the voxels in each iteration separately starting from voxels at the tumour periphery.

A better option is to use classifiers that account for neighbourhood interdependencies in the training data, given a properly selected feature set. In order to illustrate the concept of neighbourhood dependencies in training data, we first describe Markov Random Fields (MRF), followed by Conditional Random Fields (CRF), and finally Support Vector Random Fields (SVRF).

MRFs provide a generative approach for modeling local dependencies. Let Y be a set of random variables, and $G(S, E)$ a graph of S vertices and E edges. Then, Y is called a MRF on S with respect to a neighbourhood N, if and only if two conditions hold: $P(Y) > 0$ and $P(y_i|y_{S-\{i\}}) = P(y_i|y_{N_i})$, where $S - \{i\}$ denotes the set difference, $y_{S-\{i\}}$ denotes random variables in $S - \{i\}$, and N_i denotes the neighbouring random variables of random variable i. The Markov property states that the conditional distribution of a variable is dependent only on its neighbours [69, 70].

CRFs are discriminative models that maximize the conditional probability of the labels given the observations $P(Y^*|X)$ where Y^* is the joint class labels, then (X, Y) is said to be a CRF if, when conditioned on Y, the random variables y_i obey the Markov property with respect to the graph: $P(y_i|X, y_{S-\{i\}}) = P(y_i|X, y_{N_i})$ [66, 69].

SVRFs are an extension of SVMs and allow the modeling of high-dimensional spatial dependencies using a CRF framework. The SVRF model has two main components: the *observation-matching* function that captures relationships between the observations and the class labels, and the *local-consistency* function that models relationships between the labels of neighbouring data points and the observations at data points [69].

Applying all three classifiers, MRF, CRF and SVRF, to our training data represents promising research directions, as these algorithms may help model more properly the dependencies between neighbouring voxels in the 'tumour' and 'non-tumour' classes such as a voxel is more likely to become tumourous given the properties of its neighbourhood.

6.2.4 Diffusion Model Improvements

Finally, improvements to the diffusion algorithm include incorporating a brain anatomy atlas or a highly accurate tissue map that would help identify "barriers" versus "highways", *i.e.*, regions where tumour never grow (*e.g.*, the petrous bone) versus regions where gliomas diffuse at a high rate (*e.g.*, the corpus collasum). It would also be important to compute the average rates of glioma invasion in the different types of brain tissue, and to account for tumour-induced pressure on CSF regions (*e.g.*, the ventricles and sulci). In addition, including a 'probabilistic tumour map' represents a promising research direction. This tumour map will consist of tumour occurrences in different brain locations (many of which may overlap). The implementation of the tumour map requires access to a significantly large dataset of a varied population of patients, *i.e.*, patients from different age categories, different tumour grades, and tumour occurrences at as many as possible brain locations. This tumour map will help estimate the probability or likelihood of glioma invasion into an unlabeled voxel when predicting diffusion for this voxel. The implementation of this map requires designing a significantly large database of glioma instances. This database can be queried to obtain tumour occurrences at a particular voxel or region in the brain. The likelihood of malignant occurrence can then be computed for this brain region.

6.3 Summary

This final chapter presented a summary of the contribution in this thesis, and suggested future research directions at the data processing level, and in the implementation of the model.

This study concludes that, though glioma diffusion modeling is a very challenging task, it is feasible to model tumour growth in the brain, with the help of supervised learning approaches. The problem of glioma diffusion modeling can be solved as a pattern recognition task where machine learning algorithms predict tumour growth in brain regions of interest based a large set of features specific to the patient, the tumour, the brain anatomy, and the tissue adjacent to the tumour. We also conclude that, in order to implement an accurate diffusion model, the feature combinations involved in the learning task must be selected in a way that would maximize the information content available to the learning algorithm. The training data should also be generated in a way that would allow a proper representation of brain regions of interest in the training. Since the features and the data have a direct role in the performance of supervised learning systems, a highly expressive feature set will lead to properly 'recognizing' patterns of glioma cell diffusion into the brain regions of interest.

Glossary of Terms

Abnormality – A subjectively defined characteristic, assigned to brain tissue regions that may appear diseased on patients' scans and that may be distinguished from healthy tissue according to their intensities and textures on the brain image.

Angiogenesis – A clinical term refering to the formation of new blood vessels from pre-existing vessels, a process essential for growth, and for the transition of tumours from a dormant to a malignant state.

Anterior – The direction towards the front of the body in an anatomical coordinate system.

Artifact – A feature which appears in an image which is not present in the imaged object.

Artificial Neural Network – A simplified emulation of the processing of information by the human brain, used in supervised learning tasks.

Astrocytoma – Intracranial tumour that originates from an astrocyte, a type of glial cell.

Axial – A tomographic imaging plane bisecting the body into top and bottom parts.

Benign – A term refering to tumours that do not invade surrounding tissue and do not metastasize to other parts of the body. Some benign tumours may progress into malignancy.

Bias field – See *Inhomogeneity field.*

Blood-brain barrier – A physical barrier between the blood vessels in the central nervous system, and most parts of the central nervous system itself. The barrier stops many substances from traveling across it.

Brain masking – The extraction of the brain from the skull on the patient's scan.

β-spline – A piecewise polynomial function that can be recursively defined.

C4.5 – A traditional method for learning decision tree classifiers.

CCI – Cross Cancer Institute, Edmonton, Alberta.

CDM – Classification-based Diffusion Model, our proposed model for tumour growth.

Central Nervous System – The brain and the spinal cord.

Cerebral cortex – A brain structure made of grey matter.

Cerebral hemisphere – Forms one half of the human brain. Hemispheres are asymmetrical with specialized functions.

Cerebrospinal fluid – A clear fluid secreted into the ventricles, and occupies the space between the skull and the cerebral cortex. Its function is to protect the brain and the spinal cord against injuries.

CH2 – See *Colin Holmes template.*

Classification – The task of assigning a class label, after learning from a finite set of examples, to unlabeled instances based on a set of training features.

Class imbalance – A problem encountered when performing supervised learning tasks, which results from the difference in the number of training instances that belong to different classes.

Clinical Target Volume – The brain volume that includes the GTV as well as tissue immediately adjacent to the GTV and that may reasonably harbour occult cancer cells.

CSF – See *Cerebrospinal fluid.*

Colin Holmes template – An average brain template of an individual who was imaged 27 times and the images were registered to the same coordinate system.

Computed Tomogrophy – A diagnostic imaging technique in which x-ray measurements from many angles are combined in an image.

Conditional Random Fields – A model that formulates Markov Random Fields (MRF) using a conditional probability distribution, *i.e.*, $p(Y|x)$ over labels Y given a particular observation sequence x (a set of interacting features and interdependent instances), rather than a joint probability over both Y and x. A Conditional Random Field is represented as an undirected graph, it labels a new observation sequence x^* by the label y^* that maximizes the conditional probability $p(y^*|x^*)$, and has the advantage of relaxing independence assumptions required by MRFs.

Contrast – An agent or substance given to the patient prior to scanning in order to enhance the visualization of particular tissue regions with specific properties, *e.g.*, it enhances leaky blood vessels in tumour regions.

Coronal – A tomographic imaging plane bisecting the body into front and back parts.

Corpus callosum – The largest white matter structure in the brain, connecting the left and right cerebral hemispheres.

CRF – See *Conditional Random Fields.*

CT – See *Computed Tomogrophy.*

CTV – See *Clinical Target Volume.*

FLAIR – Fluid Attenuated Inversion Recovery, an MR imaging technique that produces images similar to T2-weighted images, but with free water suppressed.

Decision Tree – A rooted graph where each node denotes a decision. Classification with Decision Trees is performed starting from the root to a leaf node that will contain a class label.

Diffusion – The process of migration and infiltration of cancer cells from a primary tumour through surrounding tissue.

DTI – Diffusion Tensor Imaging, a method for imaging the white fibres in the brain using the diffusivity of water. DTI is a development of diffusion-weighted imaging.

DWI – Diffusion-Weighted Imaging, a specific MRI modality that produces in vivo images of biological tissues weighted with the local microstructural characteristics of water diffusion.

Edema – Swelling of any organ or tissue due to the accumulation of excess fluid.

Entropy – An information content measure that considers the likelihoods of individual events occuring.

Feature selection – A method of weighting and selecting the features such that only the most relevant subset of a feature set is used.

GBM – Glioblastoma Multiforme, the most common and aggressive, grade IV glioma.

Glial cell – A major cell type, in the Central Nervous System, that functions as supporting cells to maintain the signalling ability of the neurons.

Glioblastoma – Refers to the most invasive, grade IV glioma.

Glioma – A primary brain tumour originating from a glial cell.

Gompertzian growth – A growth model that views the tumour as a population of two cell classes, proliferating versus inactive cells, and describes the growth as a process in which these two classes of cells interact.

Grey matter – A category of brain tissue with many nerve cell bodies and few myelinated fibres.

GTV – See *Gross Tumour Volume.*

Gross Tumour Volume – The tumour volume detected on the patient's scan, subject to radiologist's judgment of *abnormal* tissue regions.

GW – A glioma diffusion model we developed based on the different motility rates of glioma cells in the grey versus the white matter.

High-grade tumour – Invasive tumours that aggressively grow in a relatively short period of time, leading to the patient's death.

Histology – The study of tissue sectioned as a thin slice, using a microscope, which helps identify the tumour grade based on the type of cells in the tumour.

ICBM – International Consortium for Brain Mapping, a project whose goal is the continuing development of a probabilistic reference system for the human brain: http://www.loni.ucla.edu/ICBM/

Information gain attribute selection – A method that evaluates the worth of an attribute in a feature set by measuring the information gain with respect to a class of instances, given a number of data examples.

Inhomogeneity field – A field that varies spatially across an image and that describes the deviation at each pixel from its corrputed intensity value.

Intensity standardization – The processing of MRI scans to reduce the variability in the intensities of similar tissue types across the image volumes of different patients.

KPS – *Karnofsky Performance Scale* is a numerical value that quantifies the health status of the patient. Higher KPS values may be associated with prolonged survival times.

Least squares – A regression method that minimizes the sum of the squared distance from the model to the training data.

LGT – See *Logistic Regression.*

Linear Regression – A statistical technique used to find the best-fitting linear relationship between a target (dependent) variable and its predictors (independent variables).

Logistic Regression – A statistical, linear regression method that predicts the proportions of a categorical target variable, given a dataset of labeled examples.

Low-grade tumour – Slow-growing, benign tumours.

Magnetic Resonance Imaging – An imaging technique based on the principles of NMR.

Malignant – A clinical term applied to aggressive tumours that lead to patient's death.

Markov Random Fields – A statistical model that takes into account dependencies in the labels of neighbouring instances, in addition to the factorized likelihood under the assumption of conditional independence over labels, *i.e.*, given a label, features are independent.

Metastasis – The spread of cancer cells from the original tumour location to other parts of the body.

Mixture model – A distribution constructured from multiple, often Gaussian, distributions.

MNI – Montreal Neurological Institute.

MNI305 – A dataset of 305 spatially registered normal individuals used in the construction of templates and prior probabilities.

MOA – Mixed Oligo-Astrocytomas, grade II gliomas that have two histologically different types of cells, astrocytes and oligodendroglia.

Modality – Refers to the imaging method used in acquiring an image volume, *e.g.*, T1-weighted versus T2-weighted image volumes.

MRI – See *Magnetic Resonance Imaging.*

Mutation – A permanent change to the genetic material of a cell.

Naïve Bayes – A classification algorithm that assumes that attribute values in a dataset are conditionally independent, given the class label, and that estimates the maximum likelihood of model parameters based on their frequencies over the training data.

NB – See *Naïve Bayes.*

NMR – *Nuclear Magnetic Resonance*, a spectroscopic technique used by scientists to elucidate chemical structure and molecular dynamics.

Neoplasm – Abnormal, disorganized growth in a tissue, forming a distinct mass that can be either benign or malignant.

Neural Networks – see *Artificial Neural Network.*

Necrosis – Death of cells in an organ or tissue, due to injury, treatment side effects, or lack of nutrients – *e.g.*, cells at the centre of large tumours become necrotic due to the lack of nutrients.

Occult cells – Invisible malignant cells that infiltrated from a primary tumour through surrounding normal tissue, and that remain undetectable with current imaging techniques.

PET – See *Positron Emission Tomography.*

Planning Target Volume – The treatment volume for irradiating tumours, which includes the CTV as well as an additional margin for setup variation, *e.g.*, for patient movement during irradiation.

Positron Emission Tomography – A nuclear medical imaging technique which produces a three dimensional image or map of functional processes in the body.

Posterior – The direction towards the back in an anatomical coordinate system.

Prediction – Refers to the output of tumour growth algorithms, which model the tumour at a future time given an initial tumour volume and some parameters that describe the tumour. In this thesis, *prediction* refers to CDM's output, which models tumour growth at a future time, based on learning a classifier, from labeled brain volumes, and using it in labeling the voxels of an unlabeled brain.

PTV – See *Planning Target Volume.*

Pulse – A short burst of energy which has a specific shape.

Radiotherapy – The medical use of radiation as part of cancer treatment to control malignant cells.

Sagittal – A tomographic imaging plane bisecting the body into left and right parts.

Series – A set of slices that range over one axis, and that are imaged using the same MRI acquisition protocol.

Scan – See *Study.*

Slice – An orthogonal view of an organ or tissue that can be visualized using imaging techniques, *e.g.*, with CT or MRI.

SPM – Statistical Parametric Mapping, the construction of spatially extended statistical processes used in testing hypotheses about functional imaging data sequences. In this thesis, we use SPM2 (the most recent version of SPM) to process the MRI brain scans.

Spatial interpolation – The estimation of unknown values in a spatial environment from known values in the same environment, *e.g.*, filling inter-slice gaps in an image volume.

Spatial registration – The spatial alignment of one or more images.

Spectroscopy – The study of the electromagnetic radiation frequencies absorbed by matter.

Spin-Lattice Relaxation – The return of the longitudinal magnitization to its equilibrium value along the +Z axis.

Spin-Lattice Relaxation Time (T1) – The time to reduce the difference between the longitudinal magnitization and its equilibrium magnetization by a factor of e.

Spin-Spin Relaxation – The return of the transverse magnitization to its equilibrium value, zero.

Spin-Spin Relaxation Time (T2) – The time to reduce the transverse magnetization by a factor of e.

Skull stripping – See *Brain masking*.

Study – A collection of MRI series of a patient, that includes axial, sagittal and coronal images taken on the same day.

Sulci – Grooves on the surface of the brain creating its characteristic appearance.

Supervised learning – A task that utilizes a set of attributes and labeled training examples to learn a model from these examples, given the correspondence between their attributes and their labels. The learned model can then be used in classifying unlabeled instances based on their attribute values.

Support Vector Machines – A set of related supervised learning methods. When used for classification, the SVM algorithm creates a hyperplane that separates the data into two classes with the maximum-margin, given labeled training examples, such that the distance from the closest examples (the margin) to the hyperplane is maximized.

SVM – See *Support Vector Machines*.

T1-weighted – A magnetic resonance image where the contrast is predominantly dependent on T1.

T2-weighted – A magnetic resonance image where the contrast is predominantly dependent on T2.

Template registration – The spatial alignment of an image volume to a template image volume.

Temporal interpolation – A method of modeling an intermediate volume between two time points, given an initial and a final volume corresponding to these time points respectively.

Tissue segmentation – The differentiation between tissue regions in an organ, *e.g.*, the separation of each of white matter, grey matter, and CSF in the brain.

Tumour – A swelling or an abnormal growth that can be either benign or malignant.

Tumour segmentation – The manual or automatic delineation of the GTV detected on the patient's scan.

UG – Uniform Growth, a model we developed and that assumes radial uniform growth around the tumour.

Ventricles – Cavities in the brain filled with cerebrospinal fluid.

Volume Imaging – Imaging which produces a three-dimensional image of an object.

Voxel – Volume element (the representation of a *pixel* in 3D).

Weighted least squares – A formulation of the Least Squares method that weights the error associated with individual training instances.

White matter – One of the two main components (or tissue types) of the Central Nervous System. It forms the bulk of the deep parts of the brain and the superficial parts of the spinal cord.

Bibliography

[1] Abaqus version 5.5. *Hibbit, Karlsson, and Sorensen, Inc.*, USA, 1995.

[2] E. C. Alvord Jr and C. M.Shaw. Neoplasms affecting the nervous system of the elderly. In S. Duckett, editor, *The Pathology of the Aging Human Nervous System*, pp 210–281. Lea and Febiger, Philadelphia, 1991.

[3] J. Ashburner, P. Neelin, D. L. Collins, A. C. Evans and K. J. Friston. Incorporating Prior Knowledge into Image Registration. *NeuroImage*, 6:344–352, 1997.

[4] J. Ashburner and K. J. Friston. Nonlinear Spatial Normalization using Basis Functions. *Human Brain Mapping*, 7(4):254–266, 1999.

[5] J. Ashburner and K. J. Friston. Voxel-Based Morphometry - The Methods. *NeuroImage*, 11(6):805–821, 2000.

[6] J. Ashburner. Another MRI Bias Correction Approach. *In 8th International Conference on Functional Mapping of the Human Brain, NeuroImage*, 16(2), 2002.

[7] J. Ashburner and K. Friston. Rigid body registration. In Frackowiak, R., Friston, K., Frith, C., Dolan, R., Price, C., Zeki, S., Ashburner, J., and Penny, W., editors, *Human Brain Function*, Chapter 2. Academic Press, 2nd edition.

[8] M. Awad and N. Bourbakis. Automated MRI Driven Image Guided Robotic Brain Surgery. *ICIIS 1999*, 126, 1999.

[9] K. Brodmann. Vergleichende Localisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellebaus. Leipzig, Barth, 1909.

[10] M. A. Brown and R. C.Semelka. *MRI: Basic Principles and Applications*. Wiley, Hoboken, NJ, USA, third edition, 2003.

[11] G. F. Brunton and T. E. Wheldon. Prediction of the complete growth pattern of human multiple myeloma from restricted initial measurements. *Cell Tissue Kinet.*, 10:591–594.

[12] M. Buckland and F. Gey. The relationship between recall and precision. *Journal of the American Society for Information Science*, 45(1):12–19, 1994.

[13] V. Bulitko, B. Draper, D. Lau, I. Levner and G. Zhang. *MR ADORE: Multi-resolution Adaptive Object Recognition* (Design Document), University of Alberta, 2002.

[14] P. K. Burgess, P. M. Kulesa, J. D. Murray, and E. C. Alvod, Jr. The interaction of growth rates and diffusion coefficients in a three-dimensional mathematical model of gliomas. *J. Neuropath. and Exp. Neurology*, 56(6):704–713, 1997.

[15] N. A. Campbell. *Biology*. Benjamin/Cummings Publishing Company, fourth edition, 1996.

[16] Y. Cao, C. I. Tsien, Z. Shen, D. S. Tatro, R. Ten Haken, M. L. Kessler, T. L. Chenevert and T. S. Lawrence Use of magnetic resonance imaging to assess blood-brain/blood-glioma barrier opening during conformal radiotherapy. *J Clin Oncol.*, 23(18):4127–4136, 2005.

[17] M. R. Chicoine and D. L. Silbergeld. Assessment of brain tumour cell motility in vivo and in vitro. *J. Neurosurg.*, 82:615–622, 1995.

[18] K. H. Cho, W. A. Hall, B. J. Gerbi, P. D. Higgins, W. A. McGuire and H. B. Clark. Single dose versus fractionated stereotactic radiotherapy for recurrent high-grade gliomas. *Int J Radiat Oncol Biol Phys.*, 45(5):1133–1141, 1999.

[19] V. F. Chong. Post treatment imaging in head and neck tumours. *Cancer Imaging*, 5(1):8–10, 2005.

[20] M. Clark, L. Hall, D. Goldgof, R. Velthuizen, F. Murtagh, and M. Silbiger. Automatic tumor segmentation using knowledge- based techniques. *IEEE Transactions on Medical Imaging*, 17:238–251, 1998.

[21] O. Clatz, P. Bondiau, H. Delingette, G. Malandain, M. Sermesant, S. K. Warfield and N. Ayache. In Silico Tumour Growth: Application to Glioblastomas. *MICCAI 2004, LNCS 3217*, 337–345, 2004.

[22] C.A. Cocosco, V. Kollokian, R. Kwan, A. C. Evan. Brainweb: online interface to a 3D MRI simulated brain database. *Neuroimage*, 5:S435, 1997.

[23] C. E. Coles, A. C. Hoole, S. V. Harden, et al. Quantitative assessment of inter-clinician variability of target volume delineation for medulloblastoma: quality assurance for the SIOP PNET 4 trial protocol. *Radiother Oncol.*, 69(2):189–194, 2003.

[24] D. L. Collins et al. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imag*, 17:463–468, 1998.

[25] S. E. Combs, R. Ahmadi, D. Schulz-Ertner, C. Thilmann, and J. Debus. Recurrent low-grade gliomas: the role of fractionated stereotactic re-irradiation. *J Neurooncol.*, 71(3):319–323, 2005.

[26] J. Cook, D. E. Woodward, P. Tracqui, and J. D. Murray. Resection of gliomas and life expectancy. *J. Neuro-Oncol.*, 24:131, 1995.

[27] S. W .Coons and D. K. Pearl. Mitosis identification in diffuse gliomas. *Cancer*, 82:1550–1555,1998.

[28] M. Costa, J. L. Boldrini, and R. C. Bassanezu. Optimal chemotherapy: a case study with drug resitance, saturation effect, and toxicity. *IMA Journal of Mathematics Applied in Medicine & Biology*, 11:45–49, 1994.

[29] G. C. Cruywagen, D. E.Woodward, P. Tracqui et al. The modeling of diffusive tumours. *J. Biol. Sys*, 3:937–945, 1995.

[30] W. J. Curran, J. Horton, J. S. Nelson et al. Recursive partitioning analysis of prognostic factors in three radiation therapy oncology group malignant glioma trials. *J Natl Cancer Inst*, 85:704–710, 1993.

[31] H. Damasio and A. R. Damasio. *Lesion Analysis in Neuropsychology.* Oxford Univ. Press, New York, 1989.

[32] F. Dhermain, R. de Crevoisier, F. Parker, et al. Role of radiotherapy in recurrent gliomas. *Bull Cancer.*, 91(11):883–889, 2004.

[33] F. Dhermain, D. Ducreux, F. Bidault, A. Bruna, F. Parker, T. Roujeau, A. Beaudre, J. P. Armand and C. Haie-Meder. Use of the functional imaging modalities in radiation therapy treatment planning in patients with glioblastoma. *Bull Cancer.*, 92(4):333-42, 2005.

[34] DICOM (Online). *Digital Imaging and Communications in Medicine*, http://medical.nema.org/.

[35] *DicomWorks* (Online). http://dicom.online.fr/

[36] R. Duda and P. Hart *Pattern Classification and Scene Analysis.* Wiley, New York, NY, USA, 1973.

[37] A. Evans, D. Collins and B. Milner. An MRI-based stereotactic atlas from 250 young normal subjects. *Society for Neuroscience Abstracts*, 18:408, 1992.

[38] A. Evans, S. Marrett, P. Neelin, D. Collins, K. Worsley,W. Dai, S. Milot, E. Meyer and D. Bub. Anatomatical mapping of functional activation in stereotactic coordinate space. *Neuroimage*, 1(1):43–53, 1992.

[39] A. Evans and D. Collins. A 305-member MRI-based stereotactic atlas for CBF activation studies. *In 40th Annual Meeting of the Society for Nuclear Medicine*, 1993.

[40] J. Folkman. Anti-angiogenesis: New concept for therapy of solid tumours. *Annals of Surgery*, 75:409–416, 1971.

[41] J. Folkman. Tumour angiogenesis: therapeutic implications. *New England Journal of Medicine*, 285:1182–1186, 1972.

[42] J. Folkman. Tumour angiogenesis factor. *Cancer Research*, 34:2109–2113, 1974.

[43] K. J. Friston, J. Ashburner, C. D. Frith, J. Poline, J. D. Heather and R. S. Frackowiak. Spatial Registration and Normalization of Images. *Human Brain Mapping*, 2:165–189, 1995.

[44] R. A. Gatenby. Models of tumor-host interaction as competing populations: implications for tumor biology and treatment. *J. Theor. Biol.*, 176(4):447-55, 1995.

[45] J. Gispert, S. Reig, J. Pascau, M. Vaquero and M. Desco. Inhomogeneity correction of magnetic resonance images by minimization of intensity overlapping. *International Conference on Image Processing*, 2:14–17, 2003.

[46] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Phil Trans of the Royal Soc.*, 182:513–585, 1825.

[47] E. M. Haacke, R. W. Brown, M. R. Thompson and R. Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design.* Wiley, USA, 1999.

[48] E. Halperin, G. Bentel, and E. Heinz. Radiation Therapy Treatment Planning in Supratentorial Glioblastoma Multiforme. *Int. J. Radiat. Oncol. Biol. Phys.*, 17:1347–1350, 1989.

[49] S. M. Haney, P. M. Thompson, T. F. Cloughesy, J. R. Alger and A. W. Toga. Tracking tumour growth rates in patients with malignant gliomas: a test of two algorithms. *American Journal of Neuroradiology*, 22:73–82, 2001.

[50] G. E. Hanks, A. L. Hanlon, T. E. Schultheiss, W. H. Pinover, B. Movsas, B. E. Epstein and M. A. Hunt. Dose escalation with 3D conformal treatment: five year outcomes, treatment optimization, and future directions. *Int J Radiat Oncol Biol Phys.*, 41(3):501–510, 1998.

[51] J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *Applied Statistics*, 28:100–108, 1979.

[52] S. Ho, Y. Kim, E. Bullitt and G. Gerig. Medical image segmentation by 3D level set evolution, Technical Report UNC Chapel Hill, 2002.

[53] S. Ho, E. Bullitt and G. Gerig. Level Set Evolution with Region Competition: Automatic 3D Segmentation of Brain Tumors. *Proc. 16th Int Conf on Pattern Recognition ICPR 2002, IEEE Computer Society*, editors: R. Kasturi, D. Laurendeau, C. Suen, 532–535, 2002.

[54] C. Holmes, R. Hoge, L. Collins, R. Woods, A. Toga and A. Evans. Enhancement of MR images using registration for signal averaging. *J Comput Assist Tomogr*, 22(2):324–333, 1998.

[55] K. Hoang-Xuan, A. Idbaih, K. Mokhtari, and M. Swanson. Towards a molecular classification of gliomas. *Bull Cancer.*, 92(4):310–316, 2005.

[56] F. Hochberg, and A. Pruitt. Assumptions in the Radiotherapy of Glioblastoma. *Neurology*, 907–911, 1980.

[57] J. P. Hornak. *The Basics of MRI* (Online), http://www.cis.rit.edu/htbooks/mri/.

[58] B. N. Joe, M. B. Fukui, C. C. Meltzer, et al. Brain tumor volume measurement: comparison of manual and semiautomated methods. *Radiology*, 212(3):811–816, 1999.

[59] A. R. Kansal, S. Torquato, G. R. Harsh, E. A. Chiocca and T. S. Deisboeck Simulated brain tumour growth dynamics using a three-dimensional cellular automaton. *J Theor Biol.*, 203:367–382, 2000.

[60] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya and K. R. K. Murthy. Improvements to Platt's SMO Algorithm for SVM Classifier Design. *Neural Computation*, 13(3):637–649, 2001.

[61] P. J. Kelly et al. Imaging-based stereotactic serial biopsies in untreated intracranial glial neoplasms. *J Neurosurg*, 66:865–874, 1987

[62] E. F. Kelly and C. Hunt. The limited value of cytoreductive surgery in elderly patients with malignant gliomas. *Neurosugery*, 34:62–67, 1994.

[63] F. W. Kreth, P. C. Warnke, R. Scheremet, and C. B. Ostertag. Surgical resection and radiation therapy versus biopsy and radiation therapy in the treatment of glioblastoma multiforme. *J. Neurosurg.*, 78:762–766, May 1993.

[64] H. Kolles, A. von Wangenheim, J. Rahmel, I. Niedermayer, and W. Feiden. Data-driven approaches to decision making in automated tumor grading. An example of astrocytoma grading. *Anal Quant Cytol Histol.*, 18(4):298–304, 1996.

[65] R. Kwan, A. C. Evans, G. B. Pike. An extensible MRI simulator for post-processing evaluation. *Proc. of the 4th Int. Conf. on visulatization in biomedical computing*, 96:135–140, 1996.

[66] J. Lafferty, F. Pereira and A. McCallum. Conditional random fields: probabilistic models for segmenting and labeling sequence data. *ICML*, 2001.

[67] E. R. Laws et al. Survival following surgery and prognostic factors for recently diagnosed malignant glioma: data from the Glioma Outcomes Project. *J Neurosurg.*, 99(3):467–73, 2003.

[68] S. le Cessie and J. van Houwelingen. Ridge Estimators in Logistic Regression. *Applied Statistics*, 41(1):191–201, 1992.

[69] C-H. Lee, R. Greiner and M. Schmidt. Support Vector Random Fields for Spatial Classification. *Proceedings of the 9th European Conference on Principals and Practices of Knolwedge Discovery in Data (PKDD 2005)*, Porto, Portugal, October 2005.

[70] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer-Verlag, Tokyo, 2001.

[71] P. E. Lonning et al. Genomics in breast cancer – therapeutic implications. *Nature Clinical Practice Oncology*, 2:26–33, 2005.

[72] S. Lu et al. Peritumoral diffusion tensor imaging of high-grade gliomas and metastatic brain tumors. *AJNR Am J Neuroradiol.*, 24(5):937–941, 2003.

[73] E. Mandonnet, J. Delattre, M. Tanguy, K. R. Swanson and A. Carpentier. Continuous growth of mean tumour diameter in a subset of grade II gliomas. *Ann Neurol*, 53:524–528, 2003.

[74] O. L. Mangasarian, R. Setiono and W. H. Wolberg. Pattern Recognition via Linear Programming: Theory and Application to Medical Diagnosis. *In Proceedings of the Workshop on Large-Scale Numerical Optimization*, SIAM, Philadelphia, PA, pages 22–31, 1989.

[75] M. Mascalchi, M. Filippi, R. Floris, C. Fonda, R. Gasparotti and N. Villari. Diffusion-weighted MR of the brain: methodology and clinical application. *Radiol Med*, 109(3):155–197, 2005.

[76] M. Marusic, Z. Bajzer, J. P. Freyer and S. Vuk-Pavlovic. Analysis of growth of multicellular tumour spheroids by mathematical models. *Cell Prolif.*, 27:73–94, 1994.

[77] Y. Matsukado, C. S. McCarthy, and J. W. Kernohan. The growth of glioblastoma multiforme (astrocytomas, grades 3 and 4) in neurosurgical practice. *J. Neurosurg.*, 18:636–644, 1961.

[78] G. P. Mazzara, R. P. Velthuizen, J. L. Pearlman, H. M. Greenberg and H. Wagner. Brain tumour target volume determination for radiation treatment planning through automated MRI segmentation. *Int J Radiat Oncol Biol Phys.*, 59(1):300–312, 2004.

[79] J. Mazziotta, A. Toga, A. Evans, et al. A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (ICBM). *Philosophical Transactions: Biological Sciences*, 356(1412):1293– 1322, 2001.

[80] MEDx (Online). *MEDICAL NUMERICS*, http://medx.sensor.com/products/medx/documentation.html.

[81] T Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[82] T Mitchell. *Machine Learning*, second edition, Graw-Hill, 2005.

[83] A. Mohamed, S. K. Kyriacou and C. Davatzikos. A statistical approach for estimating brain tumour induced deformation. *Mathematical Methods in Biomedical Image Analysis*, 52–59, 2001.

[84] J. D. Murray. *Mathematical Biology.* Heidelburg: Springer-Verlag, 1993.

[85] J. D. Murray. *Mathematical Biology. II: Spatial models and biomedical applications.*, third edition, Springer-Verlag, New York, 2003.

[86] A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *In Advances in Neural Information Processing Systems 14*, 2002.

[87] M. L. Peterson et al. The use of machine learning and causal inference methodology to interpret the association between drug-resistance mutations and treatment response. *Antivir Ther.*, 10, Suppl 1:S59 (abstract no. 54), 2005.

[88] J. C. Platt. Fast Training of Support Vector Machines using Sequential Minimal Optimization. *Advances in Kernel Methods - Support Vector Learning*, B. Schoelkopf, C. Burges, and A. Smola, eds., MIT Press, 1998.

[89] J. C. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, A. Smola, P. Bartlett, B. Schoelkopf, and D. Schuurmans, eds., Cambridge, MA: MIT Press, 1999.

[90] S. J. Price, N. G. Burnet, T. Donovan, et al. Diffusion tensor imaging of brain tumours at 3T: a potential tool for assessing white matter tract invasion? *Clinical Radiology*, 58:455–462, 2003.

[91] J. R. Quinlan. *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[92] F. J. Richards. A flexible growth function for empirical use. *J of Exper Bot*, 10:290–300, 1959.

[93] D. Shattuck, S. Sandor-Leahy, K. Schaper, D. Rottenberg and R. Leahy. Magnetic resonance image tissue classification using a partial volume model. *Neuroimage*, 13(5):856–876,2001.

[94] W. Schlegel, et al. Systems and mechanical tools for stereotactical guided conformation therapy with linear accelerators. *Int. J. Radiat. Oncol. Biol. Phys.*, 24:781-787, 1992.

[95] M. Schmidt. *Automatic Brain Tumour Segmentation.* MSc Thesis, University of Alberta, 2005.

[96] D. L. Silbergeld, R. C. Rostomily, and E. C. Alvord, Jr. The cause of death in patients with glioblastoma is multifactorial: Clinical factors and autopsy findings in 117 cases of supratentorial glioblastoma in adults. *J.Neuro-Oncol.*, 10:179–185, 1991.

[97] A. Simmons, P. Tofts, G. Barker, and S. Arridge. Sources of intensity non-uniformity in spin echo images at 1.5 t. *Magn Reson Med*, 32(1):121–128, 1994.

[98] J. G. Sled and G. B. Pike. Understanding intensity nonuniformity in MRI. *in Proc. Medical Image Computing Computer-Assisted Intervention – MICCAI 98* (Lecture Notes in Computer Science), Berlin, Germany: Springer-Verlag, 1496:614–622, 1998.

[99] S. Smith and J. Bradly. SUSAN – a new approach to low-level image processing. *Int. Journal of Computer Vision*, 23(1):45–78, 1997.

[100] S. M. Smith et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(S1):208–219, 2004.

[101] S. W. Song, G. N. Fuller, H. Zheng, and W. Zhang. Inactivation of the invasion inhibitory gene IIp45 by alternative splicing in gliomas. *Cancer Res.*, 65(9):3562–3567, 2005.

[102] A. G. Sorensen, et al. Comparison of diameter and perimeter methods for tumour volume calculation. *J Clin Oncol.*, 19(2):551–557, 2001.

[103] S. Draghici and R. B. Potter. Predicting HIV drug resistance with neural networks. *Bioinformatics*, 19(1):98–107, 2003.

[104] SPM (Online). Statistical parametric mapping, http://www.fil.ion.bpmf.ac.uk/spm/.

[105] W. N. Street, O. L. Mangasarian and W. H. Wolberg. An inductive learning approach to prognostic prediction. *In A. Prieditis and S. Russell, editors, Proceedings of the Twelfth International Conference on Machine Learning*, pages 522–530, Morgan Kaufmann, San Francisco, 1995.

[106] W. N. Street. A neural network model for prognostic prediction. *In J. Shavlik, editor, Proceedings of the Fifteenth International Conference on Machine Learning*, pages 540–546, Morgan Kaufmann, San Francisco, 1998.

[107] Student. The Probable Error of a Mean. *Biometrika*, 6:1–25, 1908.

[108] G. W. Swan. Tumour growth models and cancer therapy. In J.R. Thomson and B.W. Brown, editors, *Cancer Modeling*, pages 91–104. Marcel Dekker, New York, 1987.

[109] K. R. Swanson. *Mathemtical Modeling of the Growth and Control of Tumours.* PhD Thesis Washington: University of Washington, 1999.

[110] K. R. Swanson, E. C. Alvord, Jr. and J. D. Murray. A quantitative model for differential motility of gliomas in grey and white matter. *Cell Prolif.*, 33:317–329, 2000.

[111] K. R. Swanson, E. C. Alvord, Jr. and J. D. Murray. Quantifying efficacy of chemotherapy of brain tumors with homogeneous and heterogeneous drug delivery. *Acta Biotheor.*, 50(4):223–237, 2002.

[112] K. R. Swanson, E. C. Alvord, Jr. and J. D. Murray. Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy. *British Journal of Cancer*, 86:14–18, 2002.

[113] K. R. Swanson, C. Bridge, J. D. Murray and E. C. Alvord, Jr. Virtual and real brain tumours: using mathematical modeling to quantify glioma growth and invasion. *Journal of the Neurological Sciences*, 216:1–10, 2003.

[114] K. R. Swanson, E. C. Alvord, Jr. and J. D. Murray. Virtual Resection of Gliomas: Effects of Location and Extent of Resection on Recurrence. *Mathematical and Computer Modelling*, 37:1177-1190, 2003.

[115] K. R. Swanson, E. C. Alvord, Jr. and J. D. Murray. Dynamics of a model for brain tumours reveals a small window for therapeutic intervention. *Discrete and continuous dynamical systems – Series B*, 4(1):289–295, 2004.

[116] M. Tabatabai, D. K. Williams and Z. Bursac. Hyperbolastic growth models: theory and application. *Theor Biol Med Model.*, 2(1):14, 2005.

[117] J. Talairach and P. Tourneaux. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging*. Thieme Medical Publishers, 1998.

[118] P. Tofts. *Quantitative MRI of the Brain: Measuring Changes Caused by Disease*. Wiley, editor, 2003.

[119] P. Tracqui, G. C. Cruywagen, D. E. Woodward et al. A mathematical model of glioma growth: the effect of chemotherapy on spatio-temporal growth. *Cell Prolif*, 28:17–31, 1995.

[120] E. Truco. Mathematical models for cellular systems: The von foerster equation, part i. *Bulletin of Mathematical Biology*, 27:283–303, 1965.

[121] E. Truco. Mathematical models for cellular systems: The von foerster equation, part ii. *Bulletin of Mathematical Biology*, 27:449–471, 1965.

[122] N. Tzourio-Mazoyer et al. Automated anatomical labelling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single subject brain. *Neuroimage*, 15:273–289, 2002.

[123] J. R. Usher and D. Henderson. Some drug-resistant models for cancer chemotherapy. part 1: Cycle-nonspecific drugs. *IMA Journal of Mathematics Applied in Medicine & Biology*, 13:99–126, 1996.

[124] P. F. Verhulst. A note on population growth. *Correspondence Mathematiques et Physiques*, 10:113–121, 1838.

[125] H. von Foerster. Some remarks on changing populations. In F. Stohlman, editor, *The Kinetics of Cellular Proliferation*, 1959

[126] K. E. Wallner, J. H. Galicich, and G. Krol. Patterns of Failure Following Treatment for Glioblastoma Multiforme and Anaplastic Astrocytoma. *Int. J. Radiat. Oncol. Biol. Phys.*, 16(6):1404–1409, 1989.

[127] S. Webb. *The Physics of Three-Dimensional Radiation Therapy Conformal Therapy, Radiosurgery and Treatment Planning*. Bristol: IOPP, 1993.

[128] S. Webb. *The Physics of Conformal Radiotherapy: Advances in Technology*. Bristol: IOPP, 1997.

[129] W. Weibull. A statistical distribution function of wide applicability. *J of Appl Mech*, 18:293–297, 1951.

[130] M. H. Werner, S. Phuphanich, and G. H. Lyman. The increasing incidence of malignant gliomas and primary central nervous system lymphoma in the elderly. *Cancer*, 76:1634–1642, 1995.

[131] I. R. Whittle. Management of primary malignant brain tumours. *J. Neurol. Neurosurg. Psychiatry*, 60:2–5, 1996.

[132] I. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[133] D. E. Woodward, J. Cook, P. Tracqui, G. C. Cruywagen, J. D. Murray, and E. C. Alvord, Jr. A mathematical model of glioma growth: the effect of extent of surgical resection. *Cell Proliferation*, 29:269–288,1996.

[134] T. Woolsey, J. Hanaway and M. Gado. *The Brain Atlas : A Visual Guide to the Human Central Nervous System* Wiley, Hoboken, NJ, USA, second edition, 2003.

[135] S. Y. Yang, K. C. Wang, B. K. Cho, Y. Y. Kim, S. Y. Lim, S. H. Park, I. H. Kim and S. K. Kim. Radiation-induced cerebellar glioblastoma at the site of a treated medulloblastoma: case report. *J Neurosurg.*, 102(4 Suppl):417–422, 2004.

[136] Y. Zhang, M. Brady and S. Smith. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Trans. on Medical Imaging*, 20(1):45–57, 2001.

[137] A. Zijdenbos, R. Forghani and A. Evans. Automatic quantification of MS lesions in 3D MRI brain data sets: Validation of INSECT. *Medical Image Computing and Computer-Assisted Intervention*, 439–448, 1998.

[138] A. Zizzari. *Methods on Tumor Recognition and Planning Target Prediction for the Radiotherapy of Cancer.* PhD Thesis, University of Magdeburg, 2004.